

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: MICHAEL PAUL Examiner #: 73476 Date: 7/20/00
 Art Unit: 1646 Phone Number 305-7038 Serial Number: 08/656,811
 Mail Box and Bldg/Room Location: CM1/10E13 Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Method for enhancing long term memory in a subject
 Inventors (please provide full names): Dusan Bratsch

Earliest Priority Filing Date: _____

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Please search SEQ IDNO: 1 CRFE
 Reverse translate SEQ IDNO: 1
 Interference and connected Database
 Thanks.

Christina Chan
 Edward Hart
 Technical Info Specialist
 STIC / Biotech
 CM1 12C14 Tel: 305-9203

STAFF USE ONLY

Searcher: _____

Searcher Phone #: _____

Searcher Location: _____

Date Searcher Picked Up: 7/25/00Date Completed: 8/9/00

Searcher Prep & Review Time: _____

Clerical Prep Time: _____

Online Time: _____

Type of Search

NA Sequence (#) 1/1

AA Sequence (#) _____

Structure (#) _____

Bibliographic _____

Litigation _____

Fulltext _____

Patent Family _____

Other _____

Vendors and cost where applicable

Dialog _____

Questel/Orbit _____

Dr.Link _____

Lexis/Nexis _____

Sequence Systems 02

WWW/Internet _____

Other (specify) _____

This Page Blank (uspto)

51 LUSD-----db

```

51  LSD-----PFLS-----EKSEMEVEPS-----PTSPA-PLIQAEHSYLS 86

```

```

QY 139 AETEEEN--HLSPDPS--PEOV-----APINLEPELTAHNTVTSIPGGL 181
Dd 87 EEPHQSGPHTAANSDSDEBEVSEKKWYLTSTERRPATIKKEPI-----TEEQPGGL 138
QY 182 GGMELASLSLFTFELDIPYNN-----DSAVSIGGAEELIGS-PLSYDVDESTISTSGP 234
Dd 139 ----VPSTVLTITTAIS-TPEKEESPMDMNAAGDSSCOTLIPKIKLPEHDEOFLINS-P 192
QY 235 SSPETSQSSIIESSPELKYISTSSIDASKRFPYSRSS-----KSKOVSYTS-D-A 284
Dd 193 KEASVDQHLHPPTPPSSHSSDEGSLSPNRLHPFSLQASHPVYRAMPREGSALSTPPL 254
QY 285 KAPKRTIR-----TPAOPVREPHVITM-----EHLDDKKDKKKLQNKAAIRYR 324
Dd 253 TAPHLQSSGPLVLTTEEBKRTLVAEGYPIPKLPLTKSEEKALKIRIKKIKNSAOSER 312
QY 325 MKKKGDAOGIGEOEDELNTKLTKLTVDDQ-----REIKYMKNUL---MEDVCKAK 373
Dd 313 RKKKEYMDSLEKVESCSSTENLTELRRKVEYLENTNRTLQDLQGLTVMGKVSRTCKLA 372
QY 374 GLO 376
Dd 373 GTQ 375

```

```

1      RESULT 2
2      US-09-005-970-2
3      : Sequence 2, Application US/09005970
4      : Patent No. 5959079
5      :
6      : GENERAL INFORMATION:
7      : APPLICANT: - Keegan, Kathleen S.
8      : TITLE OF INVENTION: No. 5959079e1 CREBA Isoform
9      : NUMBER OF SEQUENCES: 10
10     : CORRESPONDENCE ADDRESS:
11     : ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
12     : STREET: 233 South Wacker Drive, 6300 Sears Tower
13     : CITY: Chicago
14     : STATE: Illinois
15     : COUNTRY: USA
16     : ZIP: 60606
17     :
18     : COMPUTER READABLE FORM:
19     : MEDIUM TYPE: Floppy disk
20     : COMPUTER: IBM PC compatible
21     : OPERATING SYSTEM: PC-DOS/MS-DOS
22     : SOFTWARE: Patent In Release #1.0, Version #1.30
23     : CURRENT APPLICATION DATA:
24     : APPLICATION NUMBER: US/09/005,970
25     : FILING DATE:
26     : CLASSIFICATION:
27     : PRIOR APPLICATION DATA:
28     : APPLICATION NUMBER: US/08/721,684
29     : FILING DATE:
30     : ATTORNEY/AGENT INFORMATION:
31     : NAME: Williams Jr., Joseph A.
32     : REGISTRATION NUMBER: 38,659
33     : REFERENCE/DOCKET NUMBER: 27866/33487
34     : TELECOMMUNICATION INFORMATION:
35     : TELEPHONE: 312-474-6300
36     : TELEFAX: 312-474-0448
37     : INFORMATION FOR SEQ ID NO: 2:
38     : SEQUENCE CHARACTERISTICS:
39     : LENGTH: 521 amino acids
40     : TYPE: amino acid
41     : TOPOLOGY: linear
42     : MOLECULE TYPE: protein
43     :
44     : IS-09-005-970-2

```

Query Match	6.8%;	Score 132;	DB 2;	Length 521;
Best Local Similarity	21.5%;	Pred. No. 0.0005;		
Matches 91;	Conservative 61;	Mismatches 151;	Indels 120;	Gaps 22

OY	19	EMPVOTGGGCGDKTSTSRHGGDESLIOPGATLTKLEPEEDVIGAEWMESSDLSGFLD	78
Dd	8	FOSVLQDMRKISEL--SEGETALMHTHFSEL-LDEFSQNVLG-----Q	50
OY	79	ALGDNHERLHPESNLLEFTSLTPDDSTVSKDILSTLTQPPQVNIPIRYASHGAEDFS	138
Dd	51	LISD-----PFLS-----EKSEMEVEPS-----PTSPA--PLIQAEHYSLS	86
OY	139	AETEFEN--LISPPDS--PROV-----AVINLEPELTLASMTVISPGL	161
Dd	87	EEPKQSQFTTAAASDSNDDEVESEKMYISTEPSTTIKEEP--TEBQPGI-	138
OY	182	GMELASELSLFTLEDFVNFN-----DSAVSGIGAEELLGS-PLSVDDVESTISFGP	234
Dd	139	----VPSTLITTAIS--TFPEKESPLDMNAGGDSQGLTPKIKLEPHEDQLANS-P	132
OY	235	SSPEFSQSIIESSEPELKYVSTISIDASKRFSYSMS-----KSKQVYTS-D-A	264
Dd	193	KEASVDOLHLPPTPSSHSDSEGLSPNPLHFPSSQAHSPVRAMPGRPSALSTBPL	252
OY	285	KAPKTR-----TPAOPVPEHYM-----EHLDKDKRKLONKNAALRYR	324
Dd	253	TAPKHLQSGPLVLTBEKERTLVAEGPIPTKLEPLTSEKALKKIRKIKNKISAOESR	312
OY	335	MKKKGGAOIGEBQDELELNTKLTKTVDDLO-----REIKYKNI-----MEDVCKAR	373
Dd	313	RKKKEVMDSLKVECSCTEHLERKKKVEYLENTNRTLLQLOLQTLVGVKSVTCCKLA	372
OY	374	GIO 376	
Dd	373	GTQ 375	

RESULT 3
US-08-319-866-8
Sequence 8, Application US/08319866
Patent No. 5929223
GENERAL INFORMATION:
APPLICANT: Tully, Timothy P.
APPLICANT: Ylin, Jerry C.
APPLICANT: Regulski, Michael
TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF GENES
TITLE OF INVENTION: ASSOCIATED WITH LONG-TERM MEMORY
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millita Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
Zip: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/319,866
FILING DATE: 7-OCT-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: CSHL94-03
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 266 amino acids

REFERENCE/DOCKET NUMBER: LUD 5455
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1142
TYPE: amino acids
STRANDEDNESS: single-stranded
TOPOLOGY: linear
US-08-845-528C-7

Query Match 6.5%; Score 124.5; DB 3; Length 1142;
Best Local Similarity 22.0%; Pred. No. 0.0083;
Matches 82; Conservative 48; Mismatches 133; Indels 109; Gaps 17;
QY 5 LWSDFQLAREWGLEMPVQTDGQFGDLKSTSRHGGDESLSLQPOGATLKLEPFEEDVLG 64
Db 616 LQEEFQSS-----LQSPVSTCS---SSTPSLSLQSPFESSQSPPEGPV--QSLHSPQSP 666
QY 65 AEWNESSDLGSFLDALGDNHERLHPFESNLEFTSLITPDDSTVSKDILSSTLQFP--- 120
Db 667 PEGMHSQSPLOSPEAPEDGDSLSPLQ-----IPOSPLGEDSLSL-LHFPQSP 715
QY 121 -----TOPVNIPLYASHGAEDF-----SAETEF-----ENHLSPPDPSPEQA 157
Db 716 EWEDLSPLHFPQPPQG-EDFOSSLQSPVSYICSSSTLSLQSPFESPOSPEGPAQ-S 773
QY 158 PVINLEPVE-----ITASHMTVISPDGLLGGMELASESLTTFELDFVFNDSAVG 207
Db 774 PL-QRPVSVSFYTYLASLQSHESQSPPE----- 803
QY 208 SIGGAELLSPLS--VDDVESTISFSGP--SSPETSQSSIIESSPELYKVIKSTSIDAS 263
Db 804 --GPAQSPLOSPPSPSTSSLSQSPVSSPSTSSLSKSSPSLQSPVISFSS 861
QY 264 KRSPSVRS--KSKQVKTSDAKAPKRTTPAQPVPEHVIMEHLDKDKRKLQ 315
Db 862 TSLSPSESSSPVDEYTSSTLLESSTLDESLESEPLETYTLDERVDELARLL- 920
QY 316 NKAAATRYRKK 327
Db 921 -----LKYQVK 927

RESULT 6
US-08-425-061-23
Sequence 23, Application US/08425061
Patent No. 5622829
GENERAL INFORMATION:
APPLICANT: KING, Mary-Claire
APPLICANT: FRIEDMAN, Lori
APPLICANT: OSTERMEYER, Beth
APPLICANT: ROWELL, Sarah
APPLICANT: LYNCH, Eric
APPLICANT: SZABO, Csilla
APPLICANT: LEE, Ming
TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESS: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/425,061
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 1363 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-425-061-23

Query Match 6.1%; Score 118.5; DB 1; Length 1363;
Best Local Similarity 19.2%; Pred. No. 0.039;
Matches 84; Conservative 70; Mismatches 135; Indels 149; Gaps 20;
QY 13 AREWGLEMPVQTDGQFGDLKST-----SRHGGDESLSLQ 47
Db 840 SRETSIEMSESLDAQI--LQNTFKVSKROSFAPFSPNGNAEECATFSAHG--SLKKQ 895
QY 48 POGATLKLEPFEEDVLGAEMWESSDLGSFLDALGDNHERLHPFESNLEFTSLITPDDST 107
Db 896 SPKVTFCEQKEEN-----QGNESNIKPVQT----- 922
QY 108 VSKDILSSTLQFP-----OPVNIPLYASHGAEDFSAETEF-----ENHLSPPD-----SP 153
Db 923 -----VNITAGFPVVGQDKPVDNAKSIKGGSRFLSSQFRGNETGLITPNKHGLLQNP 977
QY 154 EOAPVNLNPELVETASHMTVI-----SPDGLLGGMELASESLT-----FT 194
Db 978 YRTPPLEPIASFVTKCKKNLLEENFEHSMSPEREMENIPSTVTSIRNNIRENVFK 1037
QY 195 ELDFVFNFD--SAVGSIGGAELLSPLSDVDDVESTISFS--GPS-----SPETSQ 241
Db 1038 EASSNINEVSGSTNEVGSINEIGS--SDENIQAEIGRNRGPKLNAMLRLGLVLQPEVYK 1095
QY 242 SSIIES--PELYK-----VISTSSIDASKRSPYSRSKSKQSVKTSDAKAPKRTTP 293
Db 1096 QSLPGSNCKHPEIKKQYEVVQTVNTD---FSPYLISDNLEQPMGSSHAS-----QV 1145
QY 294 AQPVEHVIMEHLDKDKRKLQ---KNAAIRYRMKKKG-----AQGIKE 337
Db 1146 CSETPDLDLDDGELKEDTSPAENDIKRESSAVFSKVQKGELSRSPPFTTHLAQGYRRG 1205
QY 338 EQELEEELNTKLTQVDDL 355
Db 1206 AKKLESSEENLSSEDEL 1223

RESULT 7
US-08-825-886-23
Sequence 23, Application US/08825886
Patent No. 5821328
GENERAL INFORMATION:
APPLICANT: KING, Mary-Claire
APPLICANT: FRIEDMAN, Lori
APPLICANT: OSTERMEYER, Beth
APPLICANT: ROWELL, Sarah
APPLICANT: LYNCH, Eric
APPLICANT: SZABO, Csilla
APPLICANT: LEE, Ming
TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN
TITLE OF INVENTION: CANCER

```
; TELECOMMUNICATION INFORMATION;
;   TELEPHONE: (415) 494-8700
;   TELEFAX: (415) 494-8771
;   TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1852 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-425-061-24

Query Match      6.1%; Score 118.5; DB 1; Length 1852;
Best Local Similarity 19.2%; Pred. No. 0.063;
Matches 84; Conservative 70; Mismatches 135; Indels 149; Gaps 20;

QY    13 AREGLEMPVQTDPGQDLKST-----SRHGGSLSLQ 47
       || :|| : || : || : || |
Db     840 SRETSIEMESELDAQY--LQNTFKVKRSFAFPSPNGNAEECATFSAHSG--SLKKQ 895

QY    48 PQAGTKLEPEEDVLGAEMWESSDGLFDALGDNHRHLHPFSNLLETFLITPDOST 107
       | :| : || : | : | : | : |
Db     896 SPKVTFECEQEEN-----QQKNESNIKPVQT----- 922

QY    108 VSKDILSTLQFTPI----QPVNIPLYASHCAEDFSATETP---ENHILSPPD-----SP 153
        :: :|| : || : || : | : | : | : |
Db     923 -----VNITAGPPVGOKDKPVDNAKCSIKGGSRFCSSQFRGNETGLITPNKHGLLPNQ 977

QY    154 EQAPVTINLEPVELTASHTMTVI-----SPDLLGMELASESUT-----ET 194
        : :| : : :| : :| :| : | : | : |
Db     978 YRIPLPIPKSVFYTKCKNKLLLENFEHSMSPREMGNIENISTVSTRNNIRENVFK 1037

QY    195 ELDFVPNFND--SAYGSIKGAEELGLSPSLDVDVESTISFS-GPS-----SPETSQ 241
         | :| : :| : :| : :| : :| : :| : :
```


;
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-425-061-16

Query Match 6.1%; Score 118.5; DB 1; Length 1863;
Best Local Similarity 19.2%; Pred. No. 0.064;
Matches 84; Conservative 70; Mismatches 135; Indels 149; Gaps 20;

QY 13 AREWGLEMPVVQTDGQFGDLKST-----SRHGGDESLSQ 47
DB 840 SRETSIEMESELDAQY--LQNTFKVSKRSQSFAPFSPNGNAEBECATFSAHSG--SLKKQ 895
QY 48 PQGATLKLEPFEEVDVGAENWESSDLGSLDALGDNHERLHPFESNLLEFTSLITPDDST 107
DB 896 SPKVTFECEQKEEN-----QCKNESNIKPVQT----- 922
QY 108 VSKDILSSTLOFFT-----QPVNIPLYASHGAEDFSAETEF---ENHLSPPD-----SP 153
DB 923 -----VNITAGFPVVGQDKPVDNAKCSIKGSRFCLSSQFRGNETGLITPNKHGLLQNP 977
QY 154 EQVAPVINLEPVELTASHMTVI-----SPDGLLGMELASESLT-----FT 194
DB 978 YRIPPLPIKSFVTKCKKLLLENFEHSMSPEREMGNENIPSTVTSIRNNIRENVFK 1037
QY 195 ELDFVNFND--SAVSGISGAEELGSLPSVDVESTISFS--GPS-----SPETSQ 241
DB 1038 EASSNNINEVGSSSTNEVGSSINEIGS--SDENIQAEIGRNGPKLNAMLRGLVLQPEYVK 1095
QY 242 SSIIESS---PELYK-----VISTSSIDASKRFSPPYSRSKSKOSVTSDAKAPRKTRTP 293
DB 1096 QSLPGSNCKHPKIKQEEYEVQTVNTD-----FSPYLISDNLEQPMGSSHAS-----QV 1145
QY 294 AQPVEHVIMEHLDDKDRKKLQN---KNAAIRYRMKKGE-----AQGIKGE 337
DB 1146 CSETPDLDLDDGEIKEDTFAENDIKESSAVFSVKQGELSRSPPFTTHLAQGYRRG 1205
QY 338 EQELEELNTKLTKKVDDL 355
DB 1206 AKKLESSEENLSSEDEEL 1223

RESULT 11

US-08-480-784-2
; Sequence 2, Application US/08480784
; Patent No. 5693473
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean W.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

;
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,784
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-480-784-2

Query Match 6.1%; Score 118.5; DB 1; Length 1863;

Best Local Similarity 19.2%; Pred. No. 0.064;
Matches 84; Conservative 70; Mismatches 135; Indels 149; Gaps 20;

QY 13 AREWGLEMPVVQTDGQFGDLKST-----SRHGGDESLSQ 47
DB 840 SRETSIEMESELDAQY--LQNTFKVSKRSQSFAPFSPNGNAEBECATFSAHSG--SLKKQ 895
QY 48 PQGATLKLEPFEEVDVGAENWESSDLGSLDALGDNHERLHPFESNLLEFTSLITPDDST 107
DB 896 SPKVTFECEQKEEN-----QCKNESNIKPVQT----- 922
QY 108 VSKDILSSTLOFFT-----QPVNIPLYASHGAEDFSAETEF---ENHLSPPD-----SP 153
DB 923 -----VNITAGFPVVGQDKPVDNAKCSIKGSRFCLSSQFRGNETGLITPNKHGLLQNP 977
QY 154 EQVAPVINLEPVELTASHMTVI-----SPDGLLGMELASESLT-----FT 194
DB 978 YRIPPLPIKSFVTKCKKLLLENFEHSMSPEREMGNENIPSTVTSIRNNIRENVFK 1037
QY 195 ELDFVNFND--SAVSGISGAEELGSLPSVDVESTISFS--GPS-----SPETSQ 241
DB 1038 EASSNNINEVGSSSTNEVGSSINEIGS--SDENIQAEIGRNGPKLNAMLRGLVLQPEYVK 1095
QY 242 SSIIESS---PELYK-----VISTSSIDASKRFSPPYSRSKSKOSVTSDAKAPRKTRTP 293
DB 1096 QSLPGSNCKHPKIKQEEYEVQTVNTD-----FSPYLISDNLEQPMGSSHAS-----QV 1145
QY 294 AQPVEHVIMEHLDDKDRKKLQN---KNAAIRYRMKKGE-----AQGIKGE 337
DB 1146 CSETPDLDLDDGEIKEDTFAENDIKESSAVFSVKQGELSRSPPFTTHLAQGYRRG 1205
QY 338 EQELEELNTKLTKKVDDL 355
DB 1206 AKKLESSEENLSSEDEEL 1223


```

; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,554B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 51A
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-483-554B-2

```

QY 294 AQPVEHVIMEHLDKDRKKLQN---KNAATRYRMKKKGE-----AQGIKGE 337
Db 1146 CSETPDDLLDDGETKEDTSAENDIKESSAVFSKSVQKGLSRSPSPFTHTHLAQGYRRG 1205
QY 338 EQELEELNTKLTKVDDL 355
Db 1206 AKLESSEENLSEDEEL 1223

RESULT 15
US-08-798-691-4
; Sequence 4, Application US/08798691
; Patent No. 5750400
; GENERAL INFORMATION:
; APPLICANT: Murphy, Patricia D.
; APPLICANT: Allen, Antonette C.
; APPLICANT: Alvares, Christopher P.
; APPLICANT: Critz, Brenda S.
; APPLICANT: Olson, Sheri J.
; APPLICANT: Schelter, Denise B.
; APPLICANT: Zeng, Bin
; TITLE OF INVENTION: Coding Sequences of the Human
; TITLE OF INVENTION: BRCAL Gene
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; STREET: -200 Perry Parkway
; CITY: Gaithersburg
; STATE: MD
; COUNTRY: USA
; ZIP: 20877
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/798,691
; FILING DATE: 12-Feb-97
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Thomas Gallegos
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: PA-0054CIP
; TELEPHONE: 301-527-2051
; TELEFAX: 301-208-6997
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; STRAIN: BRCAL
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 17
; MAP POSITION: 17q21
US-08-798-691-4

Query Match 6.1%; Score 118.5; DB 1; Length 1863;
Best Local Similarity 19.2%; Pred. No. 0.064;
Matches 84; Conservative 70; Mismatches 135; Indels 149; Gaps 20;

QY 13 AREWGLEMPVYVDGQFGDLKST-----SRHGGDESLSLQ 47
Db 840 SRETSIEMESELDAQY--LQNTFKVSKRQSFALFNPNGNAEECATFSAHSG--SLKKQ 895
QY 48 PQGATLKLEPFEDVLGAENWESSDLGSLDALGDNHERLHPFESNLLEFTSLITPDST 107

Db 896 SPKVTFECEQKEEN-----QGNESNIKPVQT----- 922
QY 108 VSKDILSSTLQFPF-----QPVNIPLYASHGAEDFSAETEF---ENHLSPPD-----SP 153
Db 923 -----VNITAGFPVVGOKDKPVDNAKCSIKGSRFCLSOFRGNETGLITPNKHGLLQNP 977
QY 154 EQAPVINLEPVELTASHMTVI-----SPDGLLGGMELASESLT-----FT 194
Db 978 YRIPPLPIKSFVTKCKKNLLEENFEHSHSPERMGNNIPSTVSTISNNIRENVFK 1037
QY 195 ELDFVNEND--SAVGSIGGAEELGSLVDVDDVESTISFS-GPS-----SPETSQ 241
Db 1038 EASSNINEVGSTNEVGSSINEIGS--SDENIQAEGLRNRGPKLNAMLRGLGVLPQPEYK 1095
QY 242 SSIIESS---PELYK-----VISTSSIDASKRFSPPYSSSKSQSVKTSDAKAPRKRTTP 293
Db 1096 QSLPGSNCKHPKIKKQEEYEVVQTVNTD---FSPYLISDNLEQPMGSSHAS-----QV 1145
QY 294 AQPVEHVIMEHLDKDRKKLQN---KNAATRYRMKKKGE-----AQGIKGE 337
Db 1146 CSETPDDLLDDGETKEDTSAENDIKESSAVFSKSVQKGLSRSPSPFTHTHLAQGYRRG 1205
QY 338 EQELEELNTKLTKVDDL 355
Db 1206 AKLESSEENLSEDEEL 1223

Search completed: August 8, 2000, 02:39:09
Job time: 11633 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2000, 23:19:16 ; Search time 25.31 Seconds

(without alignments)
354.683 Million cell updates/sec

Title: US-08-656-811A-1

Perfect score: 1928

Sequence: 1 MELDLWSEDFQLAREWGLEM.....KYNKNLMEDVCKAKGIQLKM 379

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1923	99.7	378	1 W41508	Aplysia CAMP-respo
2	132	6.8	521	1 W37934	Mouse CAMP regulat
3	128.5	6.7	266	1 R91295	Drosophila dCREB1
4	124.5	6.5	1142	1 R81346	Tumour rejection a
5	121.5	6.3	1863	1 R81540	BRCA1 mutant from
6	120.5	6.2	654	1 R13950	E2Aalpha protein.
7	120.5	6.2	1863	1 R81514	BRCA1 mutant from
8	120.5	6.2	1863	1 R81543	BRCA1 mutant from
9	119.5	6.2	1503	1 W48845	Human receptor tyr
10	118.5	6.1	1261	1 R81523	BRCA1 mutant from
11	118.5	6.1	1363	1 R81525	BRCA1 mutant from
12	118.5	6.1	1363	1 R99447	BRCA1 allele #582
13	118.5	6.1	1363	1 W10010	Protein encoded by
14	118.5	6.1	1363	1 W79896	Tumourigenic BRCA1
15	118.5	6.1	1442	1 R81526	BRCA1 mutant from
16	118.5	6.1	1540	1 R81483	BRCA1 mutant from
17	118.5	6.1	1540	1 R81527	BRCA1 mutant from
18	118.5	6.1	1619	1 R81528	BRCA1 mutant from
19	118.5	6.1	1669	1 R81530	BRCA1 mutant from
20	118.5	6.1	1828	1 R81531	BRCA1 mutant from
21	118.5	6.1	1852	1 R99448	BRCA1 allele #77 t
22	118.5	6.1	1852	1 W10011	Protein encoded by
23	118.5	6.1	1852	1 W79897	Tumourigenic BRCA1
24	118.5	6.1	1863	1 R76641	BRCA1 protein. Det
25	118.5	6.1	1863	1 R91208	BRCA1, breast and
26	118.5	6.1	1863	1 R81481	BRCA1, New nucleic
27	118.5	6.1	1863	1 R81485	BRCA1 mutant from
28	118.5	6.1	1863	1 R81486	BRCA1 mutant from
29	118.5	6.1	1863	1 R81490	BRCA1 mutant from
30	118.5	6.1	1863	1 R81497	BRCA1 mutant from
31	118.5	6.1	1863	1 R81500	BRCA1 mutant from
32	118.5	6.1	1863	1 R81524	BRCA1 mutant from
33	118.5	6.1	1863	1 R81529	BRCA1 mutant from

ALIGNMENTS

RESULT 1

W41508

ID W41508 standard; Protein; 378 AA.

AC W41508;

DT 22-JUN-1998 (first entry)

DE Aplysia CAMP-response element binding protein 2.

KW ApCREB-2; CAMP-response element binding protein-2; snail;

KW memory loss; Alzheimer's disease; amnesia; ischaemia; head trauma;

KW neuronal injury; Parkinson's disease; senility; therapy.

OS Aplysia californica.

FH Key Location/Qualifiers

FT Region 73..108

FT /note= "leucine heptad repeat region"

FT Modified_site 150..153

FT /note= "MAP kinase phosphorylation site"

FT Modified_site 234..237

FT /note= "MAP kinase phosphorylation site"

FT Modified_site 271..174

FT /label= "protein kinase C phosphorylation site"

FT Domain 306..378

FT /note= "basic region/leucine zipper domain,

FT interacts with ApC/EBP"

PN WO9746257-A1.

PD 11-DEC-1997.

PF 03-JUN-1997; U09438

PR 03-JUN-1996; US-656811.

PA (UYCO) UNIV COLUMBIA NEW YORK.

PI Bartsch D, Ghirardi M, Kandel ER;

DR WPI: 98-051903/05.

DR N-PSDB: V04079.

PT Enhancing long-term memory in subjects whose cAMP-responsive gene is

repressed - used to treat long-term memory defects, e.g. age-related

memory loss, Alzheimer's disease

PT Example 2; Page 73-74; 100pp; English.

PS This protein comprises the CAMP-response element binding protein-2,

i.e. ApCREB-2, of the marine snail Aplysia. ApCREB-2 is a

transcription factor and repressor of long-term facilitation in

Aplysia neurons. It is constitutively expressed in sensory

neurons. ApCREB-2 is a homologue of human CREB-2 and mouse ATF-4.

The amino acid sequence of ApCREB-2 was deduced from a polynucleotide

sequence (see V04079) obtained from 2 clones isolated from an Aplysia

central nervous system cDNA library. The invention provides a method

of enhancing long-term memory in a subject whose cAMP-responsive gene

expression is repressed due to binding of a CREB-2 to a protein or

DNA associated with CAMP-responsive gene expression, or both. The

method involves administering to the subject a compound capable of

interfering with such binding so as to derepress CAMP-responsive gene

expression and thereby enhance long-term memory. Such compounds

include anti-CREB-2 antibodies or a compound capable of altering

phosphorylation of CREB-2. The method is used to treat e.g.

age-related memory loss, Alzheimer's disease, amnesia, ischaemia,

shock, head trauma, neuronal injury, toxicity or degradation,

CC Parkinson's disease or senility (claimed).

CC Sequence 378 AA;

SQ

34 118.5 6.1 1863 1 R81532 BRCA1 mutant from
35 118.5 6.1 1863 1 R81533 BRCA1 mutant from
36 118.5 6.1 1863 1 R81534 BRCA1 mutant from
37 118.5 6.1 1863 1 R81535 BRCA1 mutant from
38 118.5 6.1 1863 1 R81505 BRCA1 mutant from
39 118.5 6.1 1863 1 R81509 BRCA1 mutant from
40 118.5 6.1 1863 1 R81537 BRCA1 mutant from
41 118.5 6.1 1863 1 R81538 BRCA1 mutant from
42 118.5 6.1 1863 1 R81542 BRCA1 mutant from
43 118.5 6.1 1863 1 R81544 BRCA1 mutant from
44 118.5 6.1 1863 1 R81545 BRCA1 mutant from
45 118.5 6.1 1863 1 R81546 BRCA1 mutant from

Query Match 99.7%; Score 1923; DB 1; Length 378;
 Best Local Similarity 100.0%; Pred. No. 8.6e-150;
 Matches 378; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MELDLASEDFOLAREWGLEMPVVTQDQFGDLKSTSRHGDESLSLOPOGATLKLEPFE 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 MELDLASEDFOLAREWGLEMPVVTQDQFGDLKSTSRHGDESLSLOPOGATLKLEPFE 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 DVLGAEMWESSDGLSGFDALGDHNERLHPFESNLLEFTSLITPDSTVSKDILSSTLQFP 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 DVLGAEMWESSDGLSGFDALGDHNERLHPFESNLLEFTSLITPDSTVSKDILSSTLQFP 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 121 TQPVNIPLYASHGAEDFSAETEFENHLSPPDSPEQVAPVINLEPVELTASHMTVISPDGL 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 TQPVNIPLYASHGAEDFSAETEFENHLSPPDSPEQVAPVINLEPVELTASHMTVISPDGL 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 181 LGMELASESLTFTELDFVNFNSAVSGISGABELLGSPSLVDVESTISFGSPSPETS 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 LGMELASESLTFTELDFVNFNSAVSGISGABELLGSPSLVDVESTISFGSPSPETS 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 241 QSSIIESSPELYKVISSIDASKRFSYRSKSKSVKTSDAKAPRKTRTPAQPVPPEH 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 QSSIIESSPELYKVISSIDASKRFSYRSKSKSVKTSDAKAPRKTRTPAQPVPPEH 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 301 VIMEHLKDKRKKLQNKNAARIYRMKKKGAEQGIKGEQEELEELNTKLTKVDDLOREIK 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 VIMEHLKDKRKKLQNKNAARIYRMKKKGAEQGIKGEQEELEELNTKLTKVDDLOREIK 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 361 YMKNLMDVCKAKGIQIK 378
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 YMKNLMDVCKAKGIQIK 378
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  
```

```

RESULT 2
W37934
ID W37934 standard; Protein; 521 AA.
AC W37934;
DT 10-AUG-1998 (first entry)
DE Mouse CAMP regulatory element binding protein.
KW Mouse CAMP regulatory element binding; mCREBa; antibody; inhibition;
KW mCREBa modulator.
OS Mus sp.
PN W09812910-A2.
PD 02-APR-1998.
PF 26-SEP-1997; U17288.
PR 27-SEP-1996; US-721684.
PA (ICOS-) ICOS CORP.
PI Keegan KS;
DR WPI; 98-230326/20.
DR N-PSDB; V29204.
PT Murine CAMP regulatory element binding protein A, mCREBa - useful
PT for systematic analysis of structure and function of mCREBa and
PT identification of those molecules with which it will react
PS Claim 1; Pages 30-31; 41pp; English.
CC This is the amino acid sequence of the mouse CAMP regulatory element
CC binding (mCREBa) protein. The products of the mCREBa such as
CC antibodies can be used for the recombinant production of the protein,
CC to identify novel genes encoding binding partner polypeptides for
CC mCREBa, for the systematic analysis of the structure and function of
CC mCREBa and identification of those molecules with which it will react,
CC to identify inhibitors of mCREBa binding to other natural binding
CC partners and to generate rodents that fail to express a functional
CC mCREBa or express a variant mCREBa, useful as models for studying the
CC activities of mCREBa and mCREBa modulators in vivo.
SQ Sequence 521 AA;
  
```

Query Match 6.8%; Score 132; DB 1; Length 521;
 Best Local Similarity 21.5%; Pred. No. 0.0074;
 Matches 91; Conservative 61; Mismatches 151; Indels 120; Gaps 22;

```

QY 19 EMPVVTQDQFGDLKSTSRHGDESLSLOPOGATLKLEPFEEDVILGAEMWESSDGLSFLD 78
   | | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | |
  
```

```

Db 8 EQSYLOWDRKLSL---SEPGETEALMYHTHFSEL-LDEFSONVLG-----Q 50
QY 79 ALGDNHERLHPFESNLLEFTSLITPDSTVSKDILSSTLQPTQPVNIPLYASHGAEDFS 138
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51 LLSD-----PFLS-----EKSESMEVEPS-----PTSPA--PLIQAEHSYLS 86
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 139 AETEFEN---HLSPDPS--PEQV-----ADVINLEPVELTASHMTVISPDGLL 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 87 EEPRTQSPFTHAATSDSFENDEEVESEKWLSTEPSATIKKEPI-----TEBQPPGL- 138
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 182 GGMELASESLTFTELDFVNFN-----DSAVSGTGGAEELLGS-PLSVDDVESTISFGP 234
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 139 ----VPVTLTITAIIS-TPEKEESPLDMNAGDSSCOTLIPKIKLEPHEVDQFLNES-P 192
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 235 SSPETSOSIIESSPELYKVISSIDASKRFSYRSK-----KSKQSVKTS-D-A 284
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 193 KEASVDQLHLPTPPSSHSSDSSEGLSPNRLHPFLSQAHSVPVRAMPSPALUSTPLL 252
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 285 KAPRKTR-----TPAQVPPEHVM-----EHLDRKKRKKLQNKNAARIYR 324
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 253 TAPHKLOGSGPLVLTETEEKRTLVAEGYPIPTKLPTKSEKALKIRKIKNKISAQESR 312
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 325 MKKKGAEQGIKGEQEELEELNTKLTKVDDLO-----REIKYMKNL-----MEDVCKAK 373
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 313 RKRKEYMDSLEKKVESCSTENLELRKKVEVLENTNRTLQLOQLQTLVMGKVSRTCKLA 372
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 374 GIQ 376
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 373 GTQ 375
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  
```

```

RESULT 3
R91295
ID R91295 standard; Protein; 266 AA.
AC R91295;
DT 09-JUL-1996 (first entry)
DE Drosophila dCREB1 protein.
KW CAMP-responsive transcriptional activator; dCREB1;
KW long-term memory; transgenic animal; Insect.
OS Drosophila sp.
FH Key Location/Qualifiers
FT domain I. .215
FT /label= Activation_domain
FT region 216..237
FT /label= Basic_region
FT /note= "contains a high proportion of Arg and Lys
FT residues"
FT domain 238..264
FT /label= Leucine_zipper
FT /note= "contains periodic Leu residues"

WO9611270-A1.
PN 18-APR-1996.
PD 06-OCT-1995; U13198.
PR 07-OCT-1994; US-319866.
PR 21-DEC-1994; US-361063.
PA (COLD-) COLD SPRING HARBOR LAB.
PI Regulski M, Tully TP, Yin JC;
DR WPI; 96-209851/21.
DR N-PSDB; T14025.
PT Novel cyclic 3',5'-adenosine monophosphate responsive
PT transcriptional activator gene - used in stimulation and enhancement
PT of longterm memory
PS Claim 32; Page 100-101; 160pp; English.
CC The Drosophila CREB1 protein (R91295) is an enhancer-specific
CC activator and a member of the CREB/ATF family. It is the product
CC of a cDNA clone (T14025) obt'd. from a Drosophila head cDNA
CC library. dCREB1 mediated transcriptional activation from
CC CAMP-responsive element-contg. reporters in the Drosophila L2 cell
CC line, but this activation was not dependent of protein kinase
CC activity. Expression of the dCREB1 gene is associated with long-term
CC memory in Drosophila. dCREB1 can be used to manipulate memory, and
CC transgenic insects expressing dCREB1 can be used to assess the effects
CC of drugs on long-term memory formation.
  
```


CC alteration in the wild type BRCA1 gene, a predisposition for breast and
 CC ovarian cancer can be diagnosed. In one method, BRCA1 mRNA isolated from
 CC a tissue sample from a subject has a probe, corresponding to a fragment
 CC of the cDNA encoding the wild type BRCA1 sequence (or an allele-specific
 CC probe for a mutation of it), added to it. The conditions allow for
 CC hybridisation of the probe to the mRNA, and any hybridisation which
 CC occurs is detected. Alternatively the BRCA1 gene in the tissue sample is
 CC isolated, and a shift in electrophoretic mobility of single stranded DNA
 CC from the sample on a non-denaturing polyacrylamide gel indicates a
 CC mutation. These methods of detection can also diagnose a lesion
 CC neoplasia associated with the BRCA1 locus. The methods may be used in
 CC gene therapy, protein replacement therapy and protein mimetics, and may
 CC be used to screen for drugs in cancer therapy.
 SQ Sequence 1863 AA;

Query Match 6.3%; Score 121.5; DB 1; Length 1863;
 Best Local Similarity 19.0%; Pred. NO. 0.36;
 Matches 83; Conservative 71; Mismatches 137; Indels 145; Gaps 19;
 QY 13 ARWGLEMPVQVQDGFGLKST-----SRHGGDESLSLQ 47
 DB 840 SRTSIEMSESLDAQY--LQNTFKVSKROSFAPFSPNGNAEBCATFSAHG--SLKKQ 895
 QY 48 PQGATLKLEPFEEDVLGAEMWESSDGLSFLDALGDNHERLHPFESNLLEFTSLITPDST 107
 DB 896 SPKVTECEQKEEN-----QGNESNIKPVQT----- 922
 QY 108 VSKDILSLTQFT-----QPVNIPLYASHGAEDFSNETEF---ENHLSPD-----SP 153
 DB 923 -----VNITAGFPVVGOKPVDNAKCSIKGSRFCLSOFRGNETGLITPNKHGLLQNP 977
 QY 154 EQVAPVNLPEVELTASHMTVPI-----SPDGLLGGMELASESUTET-----ELDFV 199
 DB 978 YRTPPLPPIKSFVTKCKNLLLENPEEHSMSPEREMGNENISTVTSIRNIRENVFK 1037
 QY 200 NFNDSSAVSGTGGABELGSLP-----SVDDVESTIFS-GPS-----SPETSQSS 243
 DB 1038 EANSNNINEVSGTNEVSGSINEISGSDENIAELGNRGPKNLMLRLGLVQLPEVYKQS 1097
 QY 244 IIESS---PELYK-----VLTSSIDASKRFSYRSRKSQSVKTSIDAKAPKTRTPAQ 295
 DB 1098 LPSGNCKHPKIKQEEVQTVNTD-----FSPYLISDNLEQPMGSSHAS-----QVCS 1147
 QY 296 PVPEHVIMEHLKDKRKLQN---KNAAIRYRMKKKE-----AOGIKGEEQ 339
 DB 1148 ETDDLLDDEGEIKEDTSAENDIKESSAVFSKSVQKELSRSPSPFTHLAQYRGA 1207
 QY 340 ELEELNTKATKVDL 355
 DB 1208 KLESSEENLSSEDEEL 1223

RESULT 6
 ID R13950 standard; Protein; 654 AA.
 AC R13950;
 DT 28-NOV-1991 (first entry)
 DE E2Aalpha protein.
 KW immunoglobulin; enhancer-binding factor; chromosomal translocation;
 KW leukemia; fusion protein.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT region 524..611
 FT /note="region capable of DNA binding and dimer
 FT formation"
 FT
 PN W09113172-A.
 PD 05-SEP-1991.
 PF 22-FEB-1991; U01168.
 PR 23-FEB-1990; US-484063.
 PA (STRD) LELAND STANFORD JR UNIV.
 PA (WHIT-) WHITEHEAD INST BIOMED RE.
 PI Cleary ML, Mellentin JD, Baltimore D, Murre C, Mccaw P;

DR WPI; 91-281484/38.
 DR N-PSDB; Q13674.
 PT Detection of t(1;19) break-point-associated genes E2A and p1 -
 PT in chromosomal translocation, and prods. useful in diagnosis and
 PT therapy of human neoplasia, esp. acute lymphoblastic leukaemia
 PS disclosure; Fig 4A; 104pp; English.
 CC An additional Glu390 is present in this protein compared with the
 CC protein encoded by the sequence of Q13670.
 CC A distinctive feature of the N-terminal domain of E2Aalpha is its
 CC high content of proline and glycine.
 CC See also Q13669-75.
 SQ Sequence 654 AA;

Query Match 6.2%; Score 120.5; DB 1; Length 654;
 Best Local Similarity 20.2%; Pred. NO. 0.092;
 Matches 83; Conservative 63; Mismatches 190; Indels 75; Gaps 13;
 QY 4 DLWSEDFQ-----LAREWGLEMPVQVQDGFGLKSTSRHGG---DESLSLQPGQATLK- 54
 DB 226 ELWSPPGQAGFGPMLGGSSPLPLPGSGPVGSSGSTFGLLHQHERMGVQLHGAEVNG 285
 QY 55 -----LEPFEEDVLGAEMW---ESSDLGSLDALGDNHERLHPFESN 93
 DB 286 GLPSASSFSAPGATYGGVSSHPTPVSGADSLILGSRGTTAGSSGDALGKALASIYSPDHS 345
 QY 94 LLEFTSLITPDDSTVSKDILSLTQFTQPVNIPLYASHGAEDFSNETEFENHLSPDSP 153
 DB 346 SNFSS---SPSTPVGSPQGLAGTSQWPRAGAPGALSPSYDGGHLGQSKIEDHL----- 397
 QY 154 EQVAPVNLPEVELTASHMTVPI-----DGLLGGMELASESUTETELDFVNFNDSAVG 207
 DB 398 DEATHVLRSHAVG-TAGDMHTLLPGHALASGFTGPMGLGRH-----AG 441
 QY 208 SIGA--EELLGSLPVDVVESTISFGSPSPSTSSSIIESSPELYK-----VISTS 258
 DB 442 LVGGSHPEDGLAGTSLMHNHAAL---PSQPCTLPD---LSRPPDSYSGILGRAGATAAAS 495
 QY 259 SIDASKRFSYPYSSKSKSVKTSIDAKAPKTRTPAQVPVPEHVIMEHLKDKRKLQN 318
 DB 496 EIKREKEDENTISAADHSEEEKELKAPARTSPDEDEDLPPPOKAKEREKRRVANN 555
 QY 319 AAIYRMKKKEAAGIKGEQELEELNKLTKTKVDLQREIKYMKLMEDV 369
 DB 556 ARERLRVRDINEAFKELGRMCQLHNLSEKPTQLLILHQAQVSVILNLEQV 606

RESULT 7
 ID R81514 standard; Protein; 1863 AA.
 AC R81514;
 DT 02-OCT-1996 (first entry)
 DE BRCA1 mutant from sample set MSK17572.
 KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
 KW antibody production; germline alteration; probe; lesion neoplasia; human;
 KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT misc_difference 1008
 FT /note="M1008I"
 FT
 PN W09605306-A2.
 PD 22-FEB-1996.
 PF 11-AUG-1995; U10202.
 PR 12-AUG-1994; US-289221.
 PR 02-SEP-1994; US-300266.
 PR 16-SEP-1994; US-308104.
 PR 29-NOV-1994; US-348824.
 PR 24-MAR-1995; US-409305.
 PR 07-JUN-1995; US-483553.
 PR 07-JUN-1995; US-480784.
 PA (MYRI-) MYRIAD GENETICS INC.
 PA (CANC-) CANCER INST.
 PA (RECH-) CENT RECH DU CHUL.

PI Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;
 DR WPI; 96-139702/14.
 DR N-PSDB; T17471.
 PT New nucleic acid and polypeptide for mutant or polymorphic BRCA1
 PT gene - for diagnosis and therapy of human breast and ovarian cancer
 PT and for diagnosing pre-disposition to these cancers
 PS Claim 1; : 218pp; English.
 PS R81483-R81497 and R81499-R81546 represent mutations of the protein
 CC encoded by the human breast and ovarian cancer predisposing gene (BRCA1)
 CC (see R81481 for wild type protein). These mutations can be used as
 CC immunogens for antibody production. The mutant BRCA1 genes encoding
 CC these sequences have at least 1 mutation or polymorphism in comparison to
 CC the wild type cDNA (see T17438 for wild type). By detecting a germline
 CC alteration in the wild type BRCA1 gene, a predisposition for breast and
 CC ovarian cancer can be diagnosed. In one method, BRCA1 mRNA isolated from
 CC a tissue sample from a subject has a probe, corresponding to a fragment
 CC of the cDNA encoding the wild type BRCA1 sequence (or an allele-specific
 CC probe for a mutation of it), added to it. The conditions allow for
 CC hybridisation of the probe to the mRNA, and any hybridisation which
 CC occurs is detected. Alternatively the BRCA1 gene in the tissue sample is
 CC isolated, and a shift in electrophoretic mobility of single stranded DNA
 CC from the sample on a non-denaturing polyacrylamide gel indicates a
 CC mutation. These methods of detection can also diagnose a lesion
 CC neoplasia associated with the BRCA1 locus. The methods may be used in
 CC gene therapy, protein replacement therapy and protein mimetics, and may
 CC be used to screen for drugs in cancer therapy.
 SQ Sequence 1863 AA;

Query Match 6.2%; Score 120.5; DB 1; Length 1863;
 Best Local Similarity 19.4%; Pred. No. 0.44;
 Matches 85; Conservative 69; Mismatches 135; Indels 149; Gaps 20;

QY 13 AREWGLEMPVQTDGQFDLTKST-----SRHGGDESLSLQ 47
 Db 840 SRETSIEMESELDAQY--LQNTFKVSKROSFAPFSPNGNAEECATFSAHSG--SLKKQ 895
 QY 48 PQGATLKEPFEEDVLGAEMWESSDLGSFLDALGDNHRLHPFESNLLEFTSLITPDOST 107
 Db 896 SPKVTFCEQKEEN-----QCKNESNIKPVQT-----922
 QY 108 VSKDILSSTLQFPT----OPVNIPLYASHGAEDFSAETEF---ENHLSPPD-----SP 153
 Db 923 -----VNITAGFPVVGOKDPVDNAKSIKGRFCLSSQFRGNETGLITPNKHLGLQNP 977
 QY 154 EQVAPVINLEPVELTASHMTV-----ISPDLGLGGMELASESLT-----FT 194
 Db 978 YRIPPLPIKSFVKTKCKNLEENFEHSPEREMGNENIPSTVTSIRNNIRENVFK 1037
 QY 195 ELDFVNFND--SAVSGTGGAEELGSLPSVDDVESTIFS--GPS-----SPETSQ 241
 Db 1038 EASSSNIINEVGSSTNEVGSSINEIGS--SDENIQAEIGLRNRPKLNAMRLGLVQLQPEVYK 1095
 QY 242 SSIIESS---PELYK-----VISTSSIDASKRSPYSRKSOSVKTSDAKAPRKTTP 293
 Db 1096 QSLPGSKCKPEIKQYEEVQTVNFD-----FSPYLIINLEOPMGSSHAS-----QV 1145
 QY 294 AQPVPHEVIMHELDKKDRKKLQN---KNAAIRYRMKKGE-----AQGIKGE 337
 Db 1146 CSETPDLDLDDGEKEDTSAENDIKESSAVFSKSVQKGLSRSPSPFTTHLAQGYRG 1205
 QY 338 EQELEELNTKLTKYVDDL 355
 Db 1206 AKKLESEENLSDEEL 1223

RESULT 8

R81543
 ID R81543 standard; Protein; 1863 AA.

AC R81543;

DT 02-OCT-1996 (first entry)

DE BRCA1 mutant from PW24.

KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;

KW antibody production; germline alteration; probe; lesion neoplasia; human;
 KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.

OS Homo sapiens.
 FH Key Location/Qualifiers
 FT misc_difference 1008 /note= "M1008I"

FT W09605306-A2.
 PN 22-FEB-1996.

PD 11-AUG-1995; U10202.

PF 12-AUG-1994; US-289221.

PR 02-SEP-1994; US-300266.

PR 16-SEP-1994; US-308104.

PR 29-NOV-1994; US-348824.

PR 24-MAR-1995; US-409305.

PR 07-JUN-1995; US-483553.

PR 07-JUN-1995; US-480784.

PA (MYRI-) MYRIAD GENETICS INC.

PA (CANC-) CANCER INST.

PA (RECH-) CENT RECH DU CHUL.

PI Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;

DR WPI; 96-139702/14.

DR N-PSDB; T17506.

PT New nucleic acid and polypeptide for mutant or polymorphic BRCA1

PT gene - for diagnosis and therapy of human breast and ovarian cancer

PT and for diagnosing pre-disposition to these cancers

PS Claim 1; : 218pp; English.

CC R81483-R81497 and R81499-R81546 represent mutations of the protein
 CC encoded by the human breast and ovarian cancer predisposing gene (BRCA1)
 CC (see R81481 for wild type protein). These mutations can be used as

CC immunogens for antibody production. The mutant BRCA1 genes encoding
 CC these sequences have at least 1 mutation or polymorphism in comparison to
 CC the wild type cDNA (see T17438 for wild type). By detecting a germline

CC alteration in the wild type BRCA1 gene, a predisposition for breast and
 CC ovarian cancer can be diagnosed. In one method, BRCA1 mRNA isolated from
 CC a tissue sample from a subject has a probe, corresponding to a fragment

CC of the cDNA encoding the wild type BRCA1 sequence (or an allele-specific
 CC probe for a mutation of it), added to it. The conditions allow for

CC hybridisation of the probe to the mRNA, and any hybridisation which
 CC occurs is detected. Alternatively the BRCA1 gene in the tissue sample is

CC isolated, and a shift in electrophoretic mobility of single stranded DNA
 CC from the sample on a non-denaturing polyacrylamide gel indicates a

CC mutation. These methods of detection can also diagnose a lesion
 CC neoplasia associated with the BRCA1 locus. The methods may be used in

CC gene therapy, protein replacement therapy and protein mimetics, and may
 CC be used to screen for drugs in cancer therapy.

CC Sequence 1863 AA;

Query Match 6.2%; Score 120.5; DB 1; Length 1863;
 Best Local Similarity 19.4%; Pred. No. 0.44;
 Matches 85; Conservative 69; Mismatches 135; Indels 149; Gaps 20;

QY 13 AREWGLEMPVQTDGQFDLTKST-----SRHGGDESLSLQ 47
 Db 840 SRETSIEMESELDAQY--LQNTFKVSKROSFAPFSPNGNAEECATFSAHSG--SLKKQ 895
 QY 48 PQGATLKEPFEEDVLGAEMWESSDLGSFLDALGDNHRLHPFESNLLEFTSLITPDOST 107
 Db 896 SPKVTFCEQKEEN-----QCKNESNIKPVQT-----922
 QY 108 VSKDILSSTLQFPT----OPVNIPLYASHGAEDFSAETEF---ENHLSPPD-----SP 153
 Db 923 -----VNITAGFPVVGOKDPVDNAKSIKGRFCLSSQFRGNETGLITPNKHLGLQNP 977
 QY 154 EQVAPVINLEPVELTASHMTV-----ISPDLGLGGMELASESLT-----FT 194
 Db 978 YRIPPLPIKSFVKTKCKNLEENFEHSPEREMGNENIPSTVTSIRNNIRENVFK 1037
 QY 195 ELDFVNFND--SAVSGTGGAEELGSLPSVDDVESTIFS--GPS-----SPETSQ 241
 Db 1038 EASSSNIINEVGSSTNEVGSSINEIGS--SDENIQAEIGLRNRPKLNAMRLGLVQLQPEVYK 1095
 QY 242 SSIIESS---PELYK-----VISTSSIDASKRSPYSRKSOSVKTSDAKAPRKTTP 293

Db 1096 QSLPGSKCKPEIKKQBYEEVQVNTD-----FSPYLIISNLEQPMGSSHAS-----QV 1145
 QY 294 AQPVPHEVIMHELDKDKRKLQN---KNAAIRYRMKKKE-----AOGIKGE 337
 Db 1146 CSETPDLLDDGKEKEDTSAENDIKESSAVFSKSVOKGELSRSPSPFTTHLAQGYRRG 1205
 QY 338 EQELEELNTLTKTKVDL 355
 Db 1206 AKKLESSEENLSSEDEL 1223

RESULT 9

W48845
 ID W48845 standard; Protein; 1503 AA.
 AC W48845;
 DT 12-OCT-1998 (first entry)
 DE Human receptor tyrosine kinase LMR2_h.
 KW Receptor tyrosine kinase; LMR2_h; human; signal transduction;
 KW cancer; neurodegenerative disorder.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Peptide 1..23
 FT /label= Sig_peptide
 FT Domain 43..69
 FT /label= TMD
 FT /note= "transmembrane domain"
 FT Modified_site 409
 FT /label= O-phosphorylated
 FT Modified_site 477
 FT /label= O-phosphorylated
 FT Modified_site 552
 FT /label= O-phosphorylated
 FT Modified_site 1032
 FT /label= O-phosphorylated
 FT Modified_site 1100
 FT /label= O-phosphorylated
 FT Modified_site 1105
 FT /label= O-phosphorylated
 FT Modified_site 1365
 FT /label= O-phosphorylated
 FT Peptide 84..98
 FT /note= "immunogen"
 FT Peptide 684..698
 FT /note= "immunogen"
 FT Peptide 1098..1113
 FT /note= "immunogen"
 FT Peptide 1488..1503
 FT /note= "immunogen"
 FT Peptide 163..166
 FT /note= "protein kinase conserved motif"
 FT Peptide 281..283
 FT /note= "protein kinase conserved motif"
 FT Peptide 329..332
 FT /note= "protein kinase conserved motif"
 PN W09822507-A2.
 PD 28-MAY-1998.
 PE 21-NOV-1997; U22526.
 PR 22-NOV-1996; US-031675.
 PA (SUGEN-) SUGEN INC.
 PI John KE. Plowman CD;
 DR WPI; 98-312419/27.
 DR N-PSDB; V32452.
 PT New isolated receptor tyrosine kinase genes - which are expressed in
 PT neuronal tissues and tumour cells, useful as targets for
 PT neurodegenerative disorders or cancers
 PS Example 1; Fig 7B; 87pp; English.
 CC This is the amino acid sequence of human LMR2-h, deduced from a
 CC LMR2-h cDNA clone (see V32452). Novel rat, human and mouse LMR1,
 CC LMR2 and LMR3 (see W48841-49) define a novel family of receptors
 CC that are structurally related to receptor tyrosine kinases (RTKs).
 CC They all share the distinct motifs that typically characterise this
 CC class of enzymes, but possess extremely short extracellular domains

CC and have C-terminal tails of unprecedented length among RTKs.
 CC Expression of LMR1 and LMR3 is highly restricted to neuronal
 CC tissues with minimal expression in other adult or embryonic organs
 CC or in human tumour cell lines. LMR2 expression is limited to adult
 CC neuronal tissues, but is also very abundantly expressed in other
 CC non-neuronal foetal tissues and in numerous tumour cell lines.
 CC Based on restricted expression of all 3 LMRs to adult neuronal
 CC tissues and the up regulation of LMR2 in a wide variety of tumour
 CC cell lines, these proteins may be critical targets for
 CC neurodegenerative disorders or cancer. Anti-LMR antibodies may be
 CC used for detecting neurodegenerative diseases or cancer
 SQ Sequence 1503 AA;

Query Match 6.2%; Score 119.5; DB 1; Length 1503;

Best Local Similarity 19.5%; Pred. NO. 0.38;

Matches 93; Conservative 73; Mismatches 151; Indels 161; Gaps 24;

QY 11 QLARENGLEMPVVTQDGQFGD-----LSTSRHGGDE 42
 Db 361 QVIRERDTKLKPKOLEQPYSDRWYEVLOFCWLSPEKRPAAEDVHRLTYLRLOSQR--DS 418
 QY 43 SLSLQPOGATLK-----LEPFEEDVLGAEMMESSDLGSLDALG----- 81
 Db 419 EVDFEQOWNALKPNTNSRDSNNAAFPILDFARDLRGEMEVLTVTETSQGLSFEYVW 478
 QY 82 -----DNHERLHPFESNLEFTSLITPDDSTVSKDILSLSTLOFP----- 120
 Db 479 EAAKHDFERSRGHLDEG--LSYTSIFYP-----VEVFESSLSDPGKQDDSGQDVP 530
 QY 121 -TQPVNIPLYASHG---AEDFSAETEFENHLS-----PP-----DSPEQVAPVINLEP 164
 Db 531 LRVPGVVPVFDAAHNLVSGDYIQLKEKSGSNLELDYPPALLTTDNDNPERTGPELS-- 587
 QY 165 VELTASHMTVISPDGLLGGMELASESLTFTLDFV--NFNDSAVGSGIGABELGSLVSVD 223
 Db 588 -QLTA-----LRSVELEESS---TDEDFOSSTDPKDSLSFG----- 620
 QY 224 DVESTISFSGPSP-----ETSOSSIIESSPELYKVIYSTSIDASKRFSYRSS---K 274
 Db 621 DLHVT---SGPSPFFNIFNDVDKSEDLPSHQKIFDLMELNGVQAD--FKPATLSSLDN 675
 QY 275 SKOSVKTS--DAKAPRKTRTPAQVPVPHVIMEHLDKRKKLQNKNAAIYRMKKKGEAQ 332
 Db 676 PKESVITGHFEKEKPKRI-FDSEPLCLSDNLHQDNFDPNLVQELSENFLF-LQEKNLK 733
 QY 333 GIKGEQOELEELNTKLTK-----VDDLQREIKYMKN-----LMEDYCKANG 374
 Db 734 GSLSSKEHINDLQTELKNAGFTeamletscRNslDteLoQfaenkPGLslLQenVstkg 791

RESULT 10

DB R81523 standard; Protein; 1261 AA.
 AC R81523;
 DT 02-OCT-1996 (first entry)
 DE BRCA1 mutant from sample set TEXAS family BC215-000.
 KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
 KW antibody production; germline alteration; probe; lesion neoplasia; human;
 KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.
 OS Homo sapiens.
 PN W09605306-A2.
 PD 22-FEB-1996.
 PF 11-AUG-1995; U10202.
 PR 12-AUG-1994; US-289221.
 PR 02-SEP-1994; US-300286.
 PR 16-SEP-1994; US-308104.
 PR 29-NOV-1994; US-348824.
 PR 24-MAR-1995; US-409305.
 PR 07-JUN-1995; US-483553.
 PR 07-JUN-1995; US-480784.
 PA (MYRI-) MYRIAD GENETICS INC.
 PA (CANC-) CANCER INST.

(RECH-) CENT RECH DU CHUL.
Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;
WPI: 96-139702/14.
N-PSDB: T17480.
New nucleic acid and polypeptide for mutant or polymorphic BRCA1
gene - for diagnosis and therapy of human breast and ovarian cancer
and for diagnosing pre-disposition to these cancers
Claim 1; 218pp; English.
R11493-R81497 and R81499-R81546 represent mutations of the protein
encoded by the human breast and ovarian cancer predisposing gene (BRCA1)
(see R81481 for wild type protein). These mutations can be used as
immunogens for antibody production. The mutant BRCA1 genes encoding
these sequences have at least 1 mutation or polymorphism in comparison
to the wild type cDNA (see T17438 for wild type). By detecting a germline
alteration in the wild type BRCA1 gene, a predisposition for breast and
ovarian cancer can be diagnosed. In one method, BRCA1 mRNA isolated from
a tissue sample from a subject has a probe, corresponding to a fragment
of the cDNA encoding the wild type BRCA1 sequence (or an allele-specific
probe for a mutation of it), added to it. The conditions allow for
hybridisation of the probe to the mRNA, and any hybridisation which
occurs is detected. Alternatively the BRCA1 gene in the tissue sample is
isolated, and a shift in electrophoretic mobility of single stranded DNA
from the sample on a non-denaturing polyacrylamide gel indicates a
mutation. These methods of detection can also diagnose a lesion
neoplasia associated with the BRCA1 locus. The methods may be used in
gene therapy, protein replacement therapy and protein mimetics, and may
be used to screen for drugs in cancer therapy.
Sequence 1261 AA;
SQ

Query Match	6.1%	Score 118.5;	DB 1;	Length 1261;
Best Local Similarity	19.2%	Pred. No. 0.36;		
Matches 84;	Conservative	70;	Mismatches 135;	
			Indels 149;	Gaps 20;

[illegible]

RESULT 11

R81525
ID R81525 standard; Protein; 1363 AA.

AC R81525;
DT 02-OCT-1996 (first entry)

DE BRCA1 mutant from sample set MONTREAL family 183.

KW	Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW	antibody production; germline alteration; probe; lesion neoplasia; human;
KW	gene therapy; protein replacement therapy; protein mimetic; BRCA1.
OS	Homo sapiens.
PN	W09605306-A2.
PD	22-FEB-1996. U10202.
PF	11-AUG-1995; US-409305.
PF	12-AUG-1994; US-289221.
PR	02-SEP-1994; US-300266.
PR	16-SEP-1994; US-308104.
PR	29-NOV-1994; US-348824.
PR	24-MAR-1995; US-409305.
PR	07-JUN-1995; US-483553.
PR	07-JUN-1995; US-480784.
PA	(MYRI-) MYRIAD GENETICS INC.
PA	(CANC-) CANCER INST.
PA	(RECH-) CENT RECH DU CHUL.
PI	Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;
DR	WPI; 96-139702/14.
DR	N-PSDB: T17482.

New nucleic acid and polypeptide for mutant or polymorphic BRCA1
 gene - for diagnosis and therapy of human breast and ovarian cancer
 PT and for diagnosing pre-disposition to these cancers
 PS Claim 1: 218pp; English.
 CC R81483-R81497 and R81499-R81546 represent mutations of the protein
 CC encoded by the human breast and ovarian cancer predisposing gene (BRCA1)
 CC (see R81481 for wild type protein). These mutations can be used as
 CC immunogens for antibody production. The mutant BRCA1 genes encoding
 CC these sequences have at least 1 mutation or polymorphism in comparison to
 CC the wild type cDNA (see T17438 for wild type). By detecting a germline
 CC alteration in the wild type BRCA1 gene, a predisposition for breast and
 CC ovarian cancer can be diagnosed. In one method, BRCA1 mRNA isolated from
 CC a tissue sample from a subject has a probe, corresponding to a fragment
 CC of the cDNA encoding the wild type BRCA1 sequence (or an allele-specific
 CC probe for a mutation of it), added to it. The conditions allow for
 CC hybridisation of the probe to the mRNA, and any hybridisation which
 CC occurs is detected. Alternatively the BRCA1 gene in the tissue sample is
 CC isolated, and a shift in electrophoretic mobility of single stranded DNA
 CC from the sample on a non-denaturing polyacrylamide gel indicates a
 CC mutation. These methods of detection can also diagnose a lesion
 CC neoplasia associated with the BRCA1 locus. The methods may be used in
 CC gene therapy, protein replacement therapy and protein mimetics, and may
 CC be used to screen for drugs in cancer therapy.
 SQ Sequence 1363 AA;

Query Match	6.18;	Score 118.5;	DB 1;	Length 1363;
Best Local Similarity	19.2%;	Pred. No. 0.4;		
Matches	84;	Conservative 70;	Mismatches 135;	Indels 149;
				Gaps 20;

[illegible]


```

QY 294 AQVPPEHVIMEHLDKDKRKLON---KNAAIRYRMKKKGE-----AQGIKE 337
Db 1146 CSETPDOLLDECEIKEDTSAENDIKESSAVFSQVKGELSRSPSTHTHLAQGYRRG 1205
QY 338 EQEELEELNTKLTKVDDL 355
Db 1206 AKKLESSEENLSEDEEL 1233

```

RESULT	12	
R99447		R99447 standard; protein; 1363 AA.
ID	AC	R99447;
DT	30-JAN-1997	(first entry)
DE	BRCA1 allele #582	translation product.
KW	BRCA1 allele; breast cancer;	ovary cancer; genetic marker;
KW	genetic susceptibility;	diagnosis; therapy; antibody.
OS	Homo sapiens.	
PN	W09633271-A2.	
PD	24-OCT-1996.	
PF	19-APR-1996;	U05621.
PR	19-APR-1995;	US-425061.
PR	(REGC) UNIV CALIFORNIA.	
PI	Friedman L, King M, Lee M,	Lynch E, Ostermeyer B;
PI	Rowel S, Szabo C;	
PI	WPI; 96-485778/48.	
PR	N-PSDB; T42031.	
PT	New mutant allele(s) of the	BRCA1 gene - useful for assessing
PT	susceptibility to inherited	breast and ovarian cancer
PS	Claim 2; Page 71-75;	85pp; English.
PS	translation products (R99437-48)	of tumorigenic alleles #5803,
CC	9601, 9815, 8403, 8203, 388,	6403, 4406, 10201, 7408, 582 and 77
CC	(T42021-32) of the BRCA1	breast cancer susceptibility gene are
CC	(with the exception of the	allele #8403 product) truncated
CC	variants of the wild-type	product. The C-terminal portions of
CC	these truncated tumorigenic	translation products can be used to
CC	generate specific binding	cpds., such as antibodies, that are used
CC	diagnostically to distinguish	non-tumorigenic wild-type and
CC	tumorigenic BRCA1 translation	products in tissue samples.
CC	Sequence	1363 AA;
CC		

[illegible]

Db 1146 CSEFPDDLLDGEIKEDTSPFANDIKESSAVFSKVKGELSRSPSPFTHTHLAGQYRRG 1205

Qy 338 EQELEEINTKLTKVDDL 355
: : | | : : : |

Db 1206 AKKLESSEENLSSEDEL 1223

RESULT 13

W10010

ID W10010 standard; Protein; 1363 AA.

AC W10010;

DE 27-AUG-1997 (first entry)

DE Protein encoded by mutant BRCA1 allele #582.

KW BRCA1; breast cancer gene; protein truncation; deletion; probe;

KW genetic marker; hybridisation assay; screening; susceptibility;

KW ovarian; prostatic; chromosome 17q.

OS Homo sapiens.

PN US5622829-A.

PF 22-APR-1997.

PF 08-DEC-1993; 163959.

PR 08-DEC-1993; US-163959.

PR 18-APR-1994; US-232535.

PR 20-OCT-1994; US-326983.

PR 19-APR-1995; US-425061.

PA (REGC) UNIV CALIFORNIA.

PI Friedman L, King M, Lee M, Lynch E, Ostermeyer B;

PI Rowell S, Szabo C;

DR WPI; 97-244387/22.

DR N-PSDB; T70074.

PT BRCA1 allele(s) - useful as genetic markers for breast, ovarian and

PT prostatic cancers

PS Example 3; Column 105-112; 63pp; English.

CC W10000-11 are proteins encoded by BRCA1 mutant alleles (T70064-75) from

CC germline DNA and breast cancer patient tumours from chromosome 17q-linked

CC families. A mutation in allele #582 has led to the deletion of TCAA in

CC exon 11 and a corresponding Stop1364Stop substitution, leading to the

CC predicted protein truncation shown in the present sequence. The nucleic

CC acids can be used as probes in hybridisation assays for screening

CC patients for susceptibility to breast, ovarian or prostatic cancer.

SQ Sequence 1363 AA.

Query Match	6.1%	Score 118.5;	DB 1;	Length 1363;
Best Local Similarity	19.2%;	Pred. No. 0.4;		
Matches	84;	Conservative	70;	Mismatches 135; Indels 149; Gaps 20;
QY	13	ARENGLEMPVQTDGQFDLKST-----SRHGDESLSLQ	47	
Db	840	SRETSIEWESELDAQY--LONTFKVKROSFAPFSPNGNAEECATFSAHSG--SUKKQ	895	
QY	48	POGATLKLEPFEEEDVLGAEMWESSDLGSFLDALGDNIERLHPFESNLFTSLITPDOST	107	
Db	896	SPKVTFEECEQKEEN-----QCKNESNIKPYVT-----	922	
QY	108	VSKDILSSTLQFPT-----QPVNIPLVASHCAEDFSAEETF---ENHLSPPD-----	SP 153	
Db	923	-----VNITAGFPVVGQKDPVDDNAKCSIKGGSRFCLSQFRGNETGLITPNKHGLLQNP	977	
QY	154	EQAPVNIPLPVELTASHMTVI-----SPDGLLGMELASESLT-----	FT 194	
Db	978	YRIPLPEIKSFYKTKCKKNLLEENFEHSPSREMGNNENIPSTVSTISRNINRENVFK	1037	
QY	195	ELDPVFNND--SAVSGTGGABEELGSPLSVDDVESTIFS-GPS-----SPTSQ	241	
Db	1038	EASSSNINEVSGSSTNEVSGSSTNETGS--SDENIQAEGLNRGPKLNAMLGLVQLQPEVYK	1095	
QY	242	SSLIES--PELYK----VISPTSSIDASKRSPSSKSKOSVKTSDAKAPKTRTP	293	
Db	1096	QSLPGSKNKEIPIKQYEEVQTVNTD-----PSPLYLSNLQEQPMGSSHAS-----	QV 1145	
QY	294	AQPVPEHIVMHLDDKDRKKLQN---KNAAIRYRMKKGE-----AQGIKGE	337	
Db	1146	CSETPDLLDDGELEKEDTSAENDIKESSAVFSGSVOKGELSRSPSPFTTHLQAYVRG	1205	

QY 338 EQELELNTKLTAKVDL 355
 Db 1206 AKKLESSENLSEDEEL 1223

RESULT 14
 W79896
 ID W79896 standard; Protein; 1363 AA.
 AC W79896;
 DE 09-DEC-1998 (first entry)
 DT Tumorigenic BRCA1 protein #582.
 KW Tumorigenic allele; ovarian; breast cancer susceptibility gene; BRCA1;
 KW identification; specific mutation; allele specific antibody; detection;
 KW binding assay; treatment.
 OS Homo sapiens.
 PN US5821328-A.
 PD 13-OCT-1998.
 PF 02-APR-1997; 825886.
 PR 08-DEC-1993; US-163959.
 PR 18-APR-1994; US-232535.
 PR 20-OCT-1994; US-326983.
 PR 19-APR-1995; US-425061.
 PR 02-APR-1997; US-825886.
 PA (REGC) UNIV CALIFORNIA.
 PI Friedman L, King M, Lee M, Lynch E, Ostermeyer B,
 PI Rowell S, Szabo C;
 PI WPI; 98-567730/48.
 DR N-PSDB; V60570.
 DR Poly(peptide(s) based on mutated BRCA1 allelic sequences - useful
 PT for identification of specific allelic mutation(s) of the gene
 PT causing breast cancer
 PS Claim 1; Columns 105-112; 62pp; English.
 CC The present sequence represents a tumorigenic protein of the breast and
 CC ovarian cancer susceptibility gene (BRCA1). Exon 11 of the present allele
 CC has a 4 bp deletion, which results in a truncated protein. Several
 CC other mutagenic alleles have also been identified (see V60560-71). The
 CC polypeptides are useful for the identification of specific mutations
 CC of BRCA1. The alleles are specific mutations of the C-terminal end of
 CC BRCA1, and can be used to compare with the translation product from a
 CC patient to identify the mutation that is causing the disease. In this
 CC respect, they can also be used to raise allele specific antibodies.
 CC They can also be used for detection purposes. The antibodies can be
 CC used in binding assays e.g. ELISA. The polypeptides can also be used
 CC for the treatment of breast cancer.
 CC Sequence 1363 AA;
 SQ

Query Match 6.1%; Score 118.5; DB 1; Length 1363;
 Best Local Similarity 19.2%; Pred. No. 0.4;
 Matches 84; Conservative 70; Mismatches 135; Indels 149; Gaps 20;

QY 13 AREWGLEMPVQDTGQFGLKST-----SRHGGDESLSQ 47
 Db 840 SRETSIEMESELDAQY--LQNTFKVSKROSFAPFNPNGNAEECATFSAHSG--SLKKQ 895

QY 48 PQGATLKLEFEEVLAEMWESSDLGSLDALGDHNLHPFESNLFTSLITPDOST 107
 Db 896 SPKVTECEQKEEN-----QGNESNKPQVT-----922

QY 108 VSKDILSSTLQFT-----QPVNIPLYASHGAEDFAETEF---ENHLSPPD-----SP 153
 Db 923 -----VNITAGFPVVGOKDPVDNAKCSIKSGSRFCLSSQFRGNETGLITPNKHGLQNP 977

QY 154 EOAPVINLPEVLTASHMTVI-----SPDGLLGGMELASESLT-----ET 194
 Db 978 YRIPPLFPPIKSFVKTKCKMLLENTEEHSMSPERMGNIPISTVTSIRNNIRENVFK 1037

QY 195 ELDFVFNFD--SAVSGIGGAELGSPLSYDDVESTISFS-GFS-----SPETSQ 241
 Db 1038 EASSSNINEVGSSTNEVGSSINEIGS--SDENIQAEIGRNRGPKLNLRLGLVQLPEVYK 1095

QY 242 SSIIESS---PELYK-----VISTSSIDASKRFSPIYSRSSKSKQSVKTSDAKAPRTRTP 293

Db 1096 QSLPGSNCKHPEIKKQYEEVVQTVNTD----FSPYLIISNLQEQPMGSSHAS-----QV 1145
 QY 294 AQPVEHVIMEHLDKDKRKKLQN---KNAAIRYRMKKGE-----AQGIKGE 337
 Db 1146 CSETPDDLLDDGEIKEDTSFAENDIKESSAVFSKVQKGLSRSPSPPTHLLAQGYRRG 1205

QY 338 EQELELNTKLTAKVDL 355
 Db 1206 AKKLESSENLSEDEEL 1223

RESULT 15
 ID R81526 standard; Protein; 1442 AA.
 AC R81526;
 DT 02-OCT-1996 (first entry)
 DE BRCA1 mutant from sample set STRANG family 1900.
 KW Cancer therapy; breast and ovarian cancer predisposing gene; Immunogen;
 KW antibody production; germline alteration; probe; lesion neoplasia; human;
 KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.
 OS Homo sapiens.
 PN W09605306-A2.
 PD 22-FEB-1996.
 PF 11-AUG-1995; U10202.
 PR 12-AUG-1994; US-289221.
 PR 02-SEP-1994; US-300266.
 PR 16-SEP-1994; US-308104.
 PR 29-NOV-1994; US-348824.
 PR 24-MAR-1995; US-409305.
 PR 07-JUN-1995; US-483553.
 PR 07-JUN-1995; US-480784.
 PA (MYRI-) MYRIAD GENETICS INC.
 PA (CANC-) CANCER INST.
 PA (RECH-) CENT RECH DU CHUL.
 PI Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;
 PI WPI; 96-139702/14.
 DR N-PSDB; TL7483.
 DR New nucleic acid and polypeptide for mutant or polymorphic BRCA1
 PT gene - for diagnosis and therapy of human breast and ovarian cancer
 PT and for diagnosing pre-disposition to these cancers
 PS Claim 1; 218pp; English.
 CC R81483-R81497 and R81499-R81546 represent mutations of the protein
 CC encoded by the human breast and ovarian cancer predisposing gene (BRCA1)
 CC (see R81481 for wild type protein). These mutations can be used as
 CC immunogens for antibody production. The mutant BRCA1 genes encoding
 CC these sequences have at least 1 mutation or polymorphism in comparison to
 CC the wild type cDNA (see TL7438 for wild type). By detecting a germline
 CC alteration in the wild type BRCA1 gene, a predisposition for breast and
 CC ovarian cancer can be diagnosed. In one method, BRCA1 mRNA isolated from
 CC a tissue sample from a subject has a probe, corresponding to a fragment
 CC of the cDNA encoding the wild type BRCA1 sequence (or an allele-specific
 CC probe for a mutation of it), added to it. The conditions allow for
 CC hybridisation of the probe to the mRNA, and any hybridisation which
 CC occurs is detected. Alternatively the BRCA1 gene in the tissue sample is
 CC isolated, and a shift in electrophoretic mobility of single stranded DNA
 CC from the sample on a non-denaturing polyacrylamide gel indicates a
 CC mutation. These methods of detection can also diagnose a lesion
 CC neoplasia associated with the BRCA1 locus. The methods may be used in
 CC gene therapy, protein replacement therapy and protein mimetics, and may
 CC be used to screen for drugs in cancer therapy.
 CC Sequence 1442 AA;
 SQ

Query Match 6.1%; Score 118.5; DB 1; Length 1442;
 Best Local Similarity 19.2%; Pred. No. 0.44;
 Matches 84; Conservative 70; Mismatches 135; Indels 149; Gaps 20;

QY 13 AREWGLEMPVQDTGQFGLKST-----SRHGGDESLSQ 47
 Db 840 SRETSIEMESELDAQY--LQNTFKVSKROSFAPFNPNGNAEECATFSAHSG--SLKKQ 895

QY 48 PQGATLKLEFEEVLAEMWESSDLGSLDALGDHNLHPFESNLFTSLITPDOST 107

```
Db 896 SPKVTFECQKEEN-----QKKNESNIKPVQT----- 922
QY 108 VSKDILSSTIQFPT-----QPVNIPLYASHCAEDFSAETEF---ENHLSPPD-----SP 153
Db 923 -----VNITAGFPVVGOKDPVDNAKCSIKGGRFCLSSQFRGNETGLITPNKHGLLQNP 977
QY 154 EQVAPVINLEPVELTASHMTVI-----SPDGLGGMELASESLT-----FT 194
Db 978 YRIPPLPPIKSFVKCKKNLLEENFEHSMSPEREMGNENIPSTVSTISRNNIRENVEK 1037
QY 195 ELDFVNFND--SAVGSIGGAELLGSLVDVDDVESTISFS--GPS-----SPETSQ 241
Db 1038 EASSNINEVGSSTNEVGSSINEIGS--SDENIQAEIGRNRGPKLNAMLRLGLVLQPEVYK 1095
QY 242 SSIIESS---PELYK-----VISTSSIDASKRFSFYSRKSOSVKTSDAKAPKTRTP 293
Db 1096 QSLPGSNCKHPEIKKQYEEVQTVNTD----FSPYLISDNLEQPMGSSHAS-----QV 1145
QY 294 AQPVEHVIMEHLDKDKRKKLQN---KNAAIRYRMKKKGE-----AQGIKE 337
Db 1146 CSETPDLDLDDGEIKEDTSAENDIKESSAVFSVKQKGLSRSPSPFTHLAQGYRRG 1205
QY 338 EQELEELNTKIKTKVDDL 355
Db 1206 AKKLESSEENLSEDEEL 1223
```

Search completed: August 8, 2000, 02:33:45
Job time: 11669 sec

OM of: US-08-656-811A-1 to: GenEmbl:* out_format : pfs
Date: Aug 8, 2000 6:43 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=frames_p2n.model -DEV=xlp
-Q/cgn2.1/uspto_spool/JUS08656811/runat_25072000_142317.17103/app_query.fasta_1.443
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -CAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DLOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=100000 -USER=US08656811.@CNG1_1_3645 -NCPU=6 -ICPU=3
-LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-08-656-811A-1
Query length: 379
Database: GenEmbl:*
Database sequences: 972840
Database length: 892348106
Search time (sec): 944.170000

score_list:

Sequence	Strd Orig	EScore	Len	Documentation
gb_in1:ACU40851	+ 1923.00	2555.45	1336	! ACU40851 Aplysia californica b21
gb_pr1:HUMCREB2A	+ 264.00	344.79	1241	! M86842 Human cAMP response ele
gb_pr1:HUMTAXREBA	+ 264.00	339.91	2015	! D90209 Human mRNA for DNA bind
gb_bt930:AC015801	+ 262.00	232.20	176297	! AC015801 Homo sapiens chrom
gb_bt930:AC021195	+ 262.00	290.44	4.4e-08	209861 ! AC021195 Homo sapiens chro
gb_pr2:HS110415	+ 261.00	295.39	2.3e-08	112460 ! AL022312 Human DNA sequen
gb_TO:MUSTRANFAC	+ 259.50	338.93	8.8e-11	1224 ! L13791 Mus musculus (clone C
gb_TO:MUSMATE4A	+ 259.50	337.64	1.0e-10	1391 ! M94087 M.musculus mATF4 (mTR
gb_TO:MMATEP4	+ 255.00	331.83	2.2e-10	1365 ! X61507 M.musculus ATF-4 gene f
gb_ov:AB013138	+ 246.50	319.82	1.0e-09	1461 ! AB013138 Gallus gallus mRNA fo
gb_pr5:AC002036	+ 231.00	238.72	2.6e-06	80920 ! AC002036 Homo sapiens Chromo
gb_TO:AB012277	+ 222.00	252.27	3.5e-08	880 ! AB012277 Mus musculus mRNA for
gb_pr2:HSU03712	+ 220.00	284.63	9.3e-08	1442 ! U03712 Human TAXREB67 pseudog
gb_pr1:AB020163	+ 188.50	242.20	2.1e-05	1506 ! AB020163 Homo sapiens mRNA for
gb_bt924:AC011452	+ 188.50	196.03	0.0080	147307 ! AC011452 Homo sapiens chrom
gb_pr4:AF101388	+ 186.00	245.44	1.4e-05	784 ! AF101388 Homo sapiens activatin
gb_bt95:AC014741	+ 184.00	200.73	0.0044	50935 ! AC014741 Drosophila melanoga
gb_in2:AC005130	+ 184.00	195.65	0.0084	84395 ! AC005130 Drosophila melanoga
gb_in2:AE003670	+ 184.00	192.32	0.0129	117404 ! AE003670 Drosophila melanoga
gb_bt918:AC007208	+ 184.00	189.54	0.0184	154690 ! AC007208 Drosophila melanoga
gb_bt915:AC021154	+ 180.50	183.41	0.0403	178857 ! AC021154 Homo sapiens chrom
gb_bt915:AC036188	+ 171.50	181.88	0.0490	63276 ! AC036188 Homo sapiens chrom
gb_bt94:AC005121	+ 171.00	177.48	0.0862	91685 ! AC005121 Drosophila melanoga
gb_TO:AB012276	+ 162.00	212.64	0.0009	850 ! AB012276 Mus musculus mRNA for
gb_bt910:AC021163	+ 154.50	148.92	3.36	175994 ! AC021163 Homo sapiens clone
gb_pr3:HUMCREBPAG	+ 145.00	179.98	0.0626	2293 ! L05913 Human cAMP responsive e
gb_pr3:HUMCREBPAB	+ 145.00	179.87	0.0835	2319 ! L05913 Human cAMP responsive e
gb_pr1:HUMCREBPA	+ 145.00	178.57	0.0750	2637 ! L05515 Homo sapiens cAMP resp
gb_pr5:AF197018	+ 143.00	186.39	0.0275	931 ! AF197018 Pan troglodytes HTLV-1
gb_pr3:HUMCREBPAD	+ 142.50	178.03	0.0804	2000 ! L05912 Human cAMP responsive e
gb_TO:RATRAF2	+ 141.00	181.10	0.0542	1209 ! M65148 Rat RAF2 mRNA, complet
gb_pr1:HSCREBP1	+ 140.50	177.32	0.0981	1647 ! X15875 Human mRNA for cAMP res
gb_TO:S76659	+ 139.50	178.35	0.0771	1302 ! S76659 cyclic AMP response ele
gb_TO:S76657	+ 139.50	176.30	0.1003	1596 ! S76657 cyclic AMP response ele
gb_TO:RNU38938	+ 139.50	174.80	0.1216	1852 ! U38938 Rattus norvegicus cAMP
gb_pr1:HSATFAL	+ 137.50	174.36	0.1287	1485 ! X57197 H.sapiens mRNA for ATF
gb_ov:GGA17724	+ 137.00	171.69	0.1812	1812 ! Y17724 Gallus gallus mRNA for
gb_ov:XU16158	+ 136.50	172.79	0.1573	1520 ! U16158 Xenopus laevis activati
gb_TO:S76655	+ 135.50	169.49	0.2403	1848 ! S76655 cyclic AMP response ele
gb_TO:AF062402	+ 134.50	154.25	1.70	7348 ! AF062402 Rattus norvegicus ver
gb_pr2:SPAC25610	+ 134.50	144.31	6.07	19724 ! 270691 S.pombe chromosome 1 c

gb_pl2:S49588 + 134.00 163.31 0.5311 2800 ! S49588 ACR1-ATF/CREB transc
gb_pl1:SCSK01 + 134.00 162.61 0.5806 3000 ! X67875 S.cerevisiae SK01 ge
gb_pl1:SCYNL167C - 134.00 161.32 0.6849 3409 ! X71443 S.cerevisiae chromos
gb_pl1:SCORFSC14 - 134.00 137.35 14.81 36813 ! X92517 S.cerevisiae PSD1,
seq_name: gb_in1:ACU40851

seq_documentation_block:

LOCUS ACU40851 1336 bp mRNA 16-MAR-1996
DEFINITION Aplysia californica bZIP transcription factor (ApCREB2) mRNA,
complete cds.
ACCESSION U40851
VERSION U40851.1 GI:1123036
KEYWORDS
SOURCE California sea hare.
ORGANISM Aplysia californica
Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia;
Anaspidea; Aplysiidae; Aplysia.
REFERENCE 1 (bases 1 to 1336)
AUTHORS Bartsch,D., Ghirardi,M., Skehel,P.A., Karl,K.A., Herder,S.P.,
Chen,M., Bailey,C.H. and Kandel,E.R.
TITLE Aplysia CREB2 represses long-term facilitation: relief of
repression converts transient facilitation into long-term
functional and structural change
JOURNAL Cell 83 (6), 979-992 (1995)
MEDLINE 96107336
REFERENCE 2 (bases 1 to 1336)
AUTHORS Bartsch,D.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-1995) Dusan Bartsch, Howard Hughes Medical
Institute, Columbia University, 722 W. 168th St., New York, NY
10032, USA

FEATURES

source 1..1336
/organism="Aplysia californica"
/db_xref="taxon:8500"
/tissue_type="central nervous system"
190..1326
/gene="ApCREB2"
190..1326
/gene="ApCREB2"
/codon_start=1
/function="bZIP transcription factor; repressor"
/product="ApCREB2"
/protein_id="AAA92437.1"
/db_xref="GI:1123037"
/translation="MEDLWSEFQLAREWGLEMPVVTQCGDLKSTSRHGGDESL
SLOPQKATLKLEFDEEDVLGAENWESSDLGSLDALGDNHERLHPFESNLEFTSLIT
PDDSTVSKDLSLTLOPQTQVNIPIYASHGAEDFSAETEFENILSPDPSQVAPVI
NLEPVELTASHMTVISPDLGLGMELASELSITFELDFVFNDSAVSGIGABELLGS
PLSVDDVVESTISFGSPSPETSSOIISESPELYKVISTSSIDASKRFSYSRSSKSK
QSVKTSDAKAPKTRTPAQPVPBHVMEHLDDKKDKLONKNAIRYMRMKKEAQOI
KGEQELEEELNKLKYVDLQLEIKTAKMLMEDVCKAKGILQK"

BASE COUNT 375 a 307 c 328 g 326 t
ORIGIN

alignment_scores:

Quality: 1923.00 Length: 378
Ratio: 5.087 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-08-656-811A-1 x ACU40851 ..
Align seg 1/1 to: ACU40851 from: 1 to: 1336
1 MetGluLeuAspLeuTrpSerGluAspPheGlnLeuAlaAraGluTrpG1 17
|||||
190 ATGGAGCTGGACCTTGGAGCCGAAGATTTCACCTGGCCGAGGATGGG 239
17 yLeuGluMetProValValGlnThrAspGlyGlnPheGlyAspLeuLyss 34
|||||

```

240  GCTGGAATGCGAGTCGTCCAGACCGCATGGCCAGTTCGGTGACCTCAAAAT 289
34  erThrSerArgHisGlyGlyAspGluSerLeuSerLeuGlnProGlnGly 50
|||||
290  CAACCAAGTCGTATGCTGGCGAGCAATCTCTAAGTTTGCAGCCCCAGGCG 339
51  AlaThrLeuLysLeuGluProPheGluGluAspValLeuGlyAlaGluTr 67
|||||
340  GCTACACTGAAGTTGGAACCTTTGAGGAAGATGCTCTTGGTGCGAGAGTG 389
67  pMetGluSerSerLeuGlySerPheLeuAspAlaLeuGlyAspAsnH 84
|||||
390  GATGAGTCGTCGCGATCTCGGCTCTTTCTGGATGCTTTGGGTGACAACC 439
84  IsGluArgLeuHisProPheGluSerAsnLeuLeuGluPheThrSerLeu 100
|||||
440  ATGAGCGGCTGCATCCGTTCGAGTCAAACTTGCCTCGAGTTTCACTTCTGT 489
101  IleThrProAspSerThrValSerLysAspIleLeuSerSerThrLe 117
|||||
490  ATCACTCTGATGATTCGACGGTGTCAAAGGACATTCCTCAGCTCAACTCT 539
117  uGluPheProThrGlnProValAsnIleProLeuTyrrAlaSerHisGlyA 134
|||||
540  TCAGTTTCCAACTCAACCAAGTGAACATCCCTTATATGCAAGTCAATGGG 589
134  laGluAspPheSerAlaGluThrGluPheGluAsnHisLeuSerProPro 150
|||||
590  CCGAAGATTCTCTGCAGAGACTGAGTTTGAGAACCACTGTCGCCTCCA 639
151  AspSerProGluGlnValAlaProValIleAsnLeuGluProValGluLe 167
|||||
640  GATTCTCGGAGCAGGTAGCCCTGTCTATAATCTAGAACCAGTTGAAC 689
167  uThrAlaSerHisMetThrValIleSerProAspGlyLeuLeuGlyGlyM 184
|||||
690  CACTCGCAGCATATGACGGTGATCTCACTGATGGCTTGTGGGTGGCA 739
184  etGluLeuAlaSerGluSerLeuThrPheThrGluLeuAspPheValAsn 200
|||||
740  TGGAACTGGCTTCAGAAAGCTTACATTTACCGAACTAGACTTGTGAAC 789
201  PheAsnAspSerAlaValGlySerIleGlyGlyAlaGluGluLeuGlu 217
|||||
790  TTCAATGACAGTGTGTGGTTCAATTTGGCGGTGCTGAAGAACTTCTTGG 839
217  ySerProLeuSerValAspValGluSerThrIleSerPheSerGlyP 234
|||||
840  CTCCCACTGTCAAGTTGATGATGTGGAAGTACAATATCATTTTTCAGGTC 889
234  roSerSerProGluThrSerGlnSerSerIleIleGluSerSerProGlu 250
|||||
890  CATGTCGCCAGAACACAGCAGCAGCAGCATCATGTAATCAAGTCCCTGAA 939
251  LeuTyrrLysValIleSerThrSerSerIleAspAlaSerLysArgPheSe 267
|||||
940  TTGTACAAAGTTATCTACCTCGTCCATTCATGTCATCTAAGCGGTTTCTC 989
267  rProTyrrSerArgSerSerLysSerLysGlnSerValIlyThrSerAsp 284
|||||
990  TCCATACCTCTCGTTCCCTCCAAGTCCAAGCAATCTGTCAAGACTTCAGACG 1039
284  laLysAlaProArgLysThrArgThrProAlaGlnProValProGluHis 300
|||||
1040  CTAGGCACCTCTGTAACAGGAGACACCGCGCGACCTGTGTCAGAACAT 1089
301  ValIleMetGluHisLeuAspLysLysAspArgLysLysLeuGlnAsnLy 317
|||||
1090  GTCATCATGGAACATTTGGACAAAAAGGACAGAAAGAAAGCTTCAGAACAA 1139
317  sAsnAlaAlaIleArgTyrrArgMetLysLysGlyGluAlaGlnGlyI 334
|||||
1140  GAATGCTGCCATTAGGTATAGGTATCAAGAAAGAGGGGAGGCTCAGGGCA 1189

```

```

334  leLysGlyGluGlnGluLeuGluGluLeuAsnThrLysLeuLysThr 350
|||||
1190  TCARAAGGAGGAGCAAGTAATAGAGAACTCAACACAAAGCTTAAGACT 1239
351  LysValAspAspLeuGlnArgGluIleLysTyrrMetLysAsnLeuMetG1 367
|||||
1240  AAGGTCGATGACTTGCAAAAGAGAAATCAAGTACATGAAAAATTTAATGGA 1289
367  uAspValCysLysAlaLysGlyIleGlnLeuLys 378
|||||
1290  AGATGTTTGCAGAGCGCAAGAGTATTACGCTTAAA 1323
seq_name: gb_pr1:HUMCREB2A

seq_documentation_block:
LOCUS HUMCREB2A 1241 bp mRNA PRI 01-NOV-1994
DEFINITION Human CAMP response element regulatory protein (CREB2) mRNA,
complete cds.
ACCESSION M86842
VERSION M86842.1 GI:181040
KEYWORDS CAMP responsive element regulatory protein.
SOURCE Homo sapiens (tissue library: of J.Leiden) adult cDNA to mRNA.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1241)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Rutheria; Primates; Catarrhini; Hominidae; Homo.
Karpinski,B.A., Morle,G.D., Huggenvik,J., Uhler,M.D. and
Leiden,J.M.
TITLE Molecular cloning of human CREB-2: an ATF/CREB transcription factor
that can negatively regulate transcription from the CAMP response
element
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (11), 4820-4824 (1992)
MEDLINE 92279218
FEATURES
source Location/Qualifiers
1..1241
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="Jurkat"
/cell_type="T lymphocyte"
/dev_stage="adult"
/tissue_lib="of J.Leiden"
/map="2q32"
108..1163
/gene="CREB2"
108..1163
/gene="CREB2"
/codon_start=1
/product="CAMP response element regulatory protein"
/protein_id="AAA52071.1"
/db_xref="GI:181041"
/translation="MTFMSFLSSEVLVDLMSFPDFPSPGLGAEESGLLDDYLEVAKHF
KPHGFSSDKAKAGSSEWLAVDGLVSPNSNKEDAFSGTDMLEKMDLKEFDLIDLGI
DDLETMPDDLTTLDLDTDLFAPLVQETNKQPQVNPVNIHLPESLTQPDQVAPFTFL
QPLPLSPGVLSSTPDHFSFSLGSEVDITEGDRKPDYTAIVAMIPCIKEEDTPSDND
SGICMSPEYLGSPOHSPSTGRSPNRSLSPGVLCGSARPKYPDPGKWAQVYKGE
KLDKLLKMEQNKTRATRYRQKRAEQEALTGECKELEKKNALERADSLAKEIQL
KDLIEVRKARKKRPP"
polyA_site 1241
/gene="CREB2"
/note="G00-128-011"
BASE COUNT 328 a 310 c 336 g 267 t
ORIGIN Chromosome 2q32.

alignment_scores:
Quality: 264.00 Length: 408
Ratio: 1.294 Gaps: 22
Percent Similarity: 50.000 Percent Identity: 28.676

alignment_block:
US-08-656-811A-1 x HUMCREB2A

```

Align seg 1/1 to: HUMCREB2A from: 1 to: 1241

```
19 GluMetProValGlnThrAspGlyGlnPheGlyAspLeuLysSer... 34
||||| : : : : : : : : : : : : : : : : : : : : : : : :
114 GAAATGAGCTTCTCAGCAGCAGGAGTGTGTGGTGGGAGCTTGATGCC 163
35 .....ThrSerArgHisGlyGlyAspGlySerLeuSerLeu..... 46
||||| : : : : : : : : : : : : : : : : : : : : : : : :
164 CPTGACCCGTCGGGTGGGTGGGAGTGAAGAACCTAGGCTCTTAGATG 213
47 .....GlnProGlnClyAlaThr..... 52
214 ATTACTGGAGGTGCCAAGCACCTTCAAACTCATGCTGCTTCCAGCGAC 263
53 .....LeuLysLeuGluProPhe..... 58
264 AAGGCTAAGCGGGCTCTCCGAATGGCTGGTGGTGGTGGTGGTGCAG 313
59 .....GluGluAspValLeuGlyAlaGluTrpMet. 68
314 TCCCTCCCAACACACAGCAGGAGGAGTCCCTCTCCGGGACAGATTGGATG 363
69 ..GluSerSerAspLeuGlySerPhe...LeuAspAlaLeuGlyAspAsn 83
||||| : : : : : : : : : : : : : : : : : : : : : : : :
364 TCGAGAAATGAGATTGAAGGAGTTCGACTTGGATGCC..... 401
84 HisGluArgLeuHisProPheGluSerAsnLeuLeuGluPheThrSerLe 100
||||| : : : : : : : : : : : : : : : : : : : : : : : :
402 .....CTGTGGGTATAGTACCT 421
100 uIleThr...ProAspAspSerThrValSerLysAsp..... 111
||||| : : : : : : : : : : : : : : : : : : : : : : : :
422 GAAACCATGTCAGATGACCTTCTGACCCAGTGGATGACACATGTGTGATC 471
112 .....IleLeuSerSerThrLeuGlnPheProThrGlnProVal 124
: : : : : : : : : : : : : : : : : : : : : : : :
472 TCTTTGGCCCCCTAGTCAGGAGACTAATAAGCAGCCCCCAGCAGCGTG 521
125 AsnIleProLeuTyrrAlaSerHisGlyAlaGluAspPheSerAlaGluTh 141
||||| : : : : : : : : : : : : : : : : : : : : : : : :
522 AAC...CCAAAT..... 530
141 rGluPheGluAsnHisLeuSerProProAspSer.....ProGluG 155
: : : : : : : : : : : : : : : : : : : : : : : :
531 .....GCCATCTC.....CCAGAAAGTTTAACAAAACCCGACC 564
155 InValAlaProValIleAsnLeuGluProValGluLeuThrAlaSerHis 171
||||| : : : : : : : : : : : : : : : : : : : : : : : :
565 AGGTTGCCCTTCCACCTTCTTACAACTCTTCCCTTTCCCCAGGGGTC 614
172 MetThrValIleSerProAspGlyLeuLeuGlyGlyMetGluLeuAlaSe 188
: : : : : : : : : : : : : : : : : : : : : : : :
615 CTGTGCC...TCCACTCCAGATCATCTCTTT...AGTTTAGAGCTGGGCAG 658
188 rGluSerLeuThrPheThrGluLeuAspPheValAsnPheAsnAspSera 205
||||| : : : : : : : : : : : : : : : : : : : : : : : :
659 TGAA...GTGGATATCATCTGAAGGAGATAGGAAGCCAGACTACACTGCTT 705
205 laValGlySerIleGlyAlaGluGluLeuLeuGlySerProLeuSer 221
||||| : : : : : : : : : : : : : : : : : : : : : : : :
706 ACGTTGCCATGATC.....CCTCAGTGC 728
222 ValAspAspValGluSerThrIleSerPheSerGlyProSerSerProGl 238
: : : : : : : : : : : : : : : : : : : : : : : :
729 ATAAAGGAGGAAGAC.....ACCCCTC 751
238 uThrSerGlnSerIleIleGluSerSerProGluLeuTyrrLysValI 255
: : : : : : : : : : : : : : : : : : : : : : : :
752 AGATAATAGTAGTCATCTGT...ATAGCCAGCAGATCCTATCTG.... 794
255 leSerThrSerSerIleAspAlaSerLysArgPheSerProTyrrSerArg 271
: : : : : : : : : : : : : : : : : : : : : : : :
```

```
795 .....GGGTCTCTCAGCACACAGCCCTCTACCAGG 824
272 SerSerLysSerLysGlnSerValLysThrSerAspAlaLysAlaProAr 288
: : : : : : : : : : : : : : : : : : : : : : : :
825 GGCTCTCCAAATAGG...AGCTCCCATCTCCAGGTGTTCTCTGTGGTC 871
288 gLysThrArgThrProAlaGlnProValProGluHisValIleMet.... 303
||||| : : : : : : : : : : : : : : : : : : : : : : : :
872 TGCCCGTCCCAAACTTACGATCCTCTGGAGAGAGATGGTAGCAGCA 921
304 .....GluHisLeuAspLysLysAspArgLysLysLeuGlnAsn 316
||||| : : : : : : : : : : : : : : : : : : : : : : : :
922 AAGTAAAGGCTGAGAACTGGATAAGACCTGAAAAAATGGAGCAAAAC 971
317 LysAsnAlaAlaIleArgTyrArgMetLysLysLysGlyGluAlaGlnG 333
: : : : : : : : : : : : : : : : : : : : : : : :
972 AAGACAGCAGCCACTAGGTACCGCAGAGAAGAGGGCGGAGCAGAGGC 1021
333 yIleLysGlyGluGluGlnLeuGluGluLeuAsnThrLysLysLeuYst 350
: : : : : : : : : : : : : : : : : : : : : : : :
1022 TCTTACTGCTGAGTGCAAAGAGCTGGAAAGAAGACGAGGCTCTAAAG 1071
350 hrLysValAspAspLeuGlnArgGluIleLysTyrMetLysAsnLeuMet 366
: : : : : : : : : : : : : : : : : : : : : : : :
1072 AGAGGCGGATTCCTGCCAAGGAGATCCAGTACCTGAAGATTGTGATA 1121
367 GluAspValCysLysAlaLysGly 374
: : : : : : : : : : : : : : : : : : : : : : : :
1122 GAAGAGTCCGCAAGCGCAAGGGG 1145
```

seq_name: gb_prl:HUMTAXREBA

seq_documentation_block:

LOCUS HUMTAXREBA 2015 bp mRNA PRI 07-FEB-1999
DEFINITION Human mRNA for DNA binding protein TAXREB67.
ACCESSION D90209
VERSION D90209.1 GI:220087
KEYWORDS CREB; DNA binding protein; TAX-responsive enhancer.
SOURCE Human fibroblast cell line GM637, cDNA to mRNA, clone pcD67.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominiidae; Homo.

REFERENCE
AUTHORS Tsujimoto,A., Nynoya,H., Morita,T., Sato,T. and Shimotohno,K.
TITLE Isolation of cDNAs for DNA-binding proteins which specifically bind
to a tax-responsive enhancer element in the long terminal repeat of
human T-cell leukemia virus type I
J. Virol. 65 (3), 1420-1426 (1991)
91140735
COMMENT Submitted (27-AUG-1990) to ddbj by:
Kunitada Shimotohno
National Cancer Center Research Institute
5-1-1 Tsukiji, Chuo-ku
Tokyo 104
Japan
Phone: 03-542-2511
Fax: 03-545-3567,
Location/Qualifiers
1. .2015
/organism="Homo sapiens"
/db_xref="taxon:9606"
882. .1937
/translation="MTEMSFLSVGLSPDPSPGLCAEESLGLDDYLVAKHF
KPHGFSSDKAKAGSEWLAVDGLVSPNSNKEADAFSCTDWMLEKMDLKEFDLALLGI
DDLETWPDLLTTLTLDCTDLFAPLVQETNKQPTVNPICHLPSLTKPKQVAPFTFL
QPLSPGLUSSTPDHSFSLSELSVDITSGDRKPDPTATVAVIPIQCIKEEDTPSDND
SGICMSPESYLSGPSHSRSPSLSPGLCGSARPKPYDPPEKNVAARVKGE
KLDKLIKMPQNKTAATRYRQKKRAEQEALTGECKELEKKNKALKERADSLAKEIQLV
KDLIEVRKARGKKRVP"

JOURNAL MEDLINE
COMMENT

FEATURES
source
CDS

BASE COUNT 450 a 534 c 609 g 422 t
ORIGIN

alignment_scores:

Quality: 264.00 Length: 408
Ratio: 1.294 Gaps: 22
Percent Similarity: 50.000 Percent Identity: 28.676

alignment_block:

US-08-656-811a-1 x HUMTAXREBA

Align seg 1/1 to: HUMTAXREBA from: 1 to: 2015

```

19 GluMetProValValGlnThrAspGlyGlnPheGlyAspLeuLysSer... 34
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
888 GAAATGACCTTCTCAGCAGCGAGGTGTGGTGGGGGACTTGATCTCCCC 937
35 .....ThrSerArgHisGlyGlyAspGluSerLeuSerLeu..... 46
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
938 CTTCCACCCGTCGGGTTGGGGGCTGAAGAAAGCCTAGTCTCTTAGATG 987
47 .....GlnProGlnGlyAlaThr..... 52
988 ATTACCTGGAGTGCGCCAAAGCAGCTTCAACCTCATGGGTTCTCCAGCGAC 1037
53 .....LeuLysLeuGluProPhe..... 58
1038 AAGGCTAAGCGGGCTCTCCGAATGGCTGTGGGTGGTGGTGGTGCAG 1087
59 .....GluGluAspValLeuGlyAlaGluTrpMet. 68
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1088 TCCCTCCCAACACAGCAAGGAGGATGCTTCTCCGGGACAGATTGGATGT 1137
69 ..GluSerSerAspLeuGlySerPhe...LeuAspAlaLeuGlyAspAsn 83
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1138 TGGAGAAATGGATTGAAGGAGTTCCGATTTGGATGCC..... 1175
84 HisGluArgLeuHisProPheGluSerAsnLeuLeuGluPheThrSerLe 100
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1176 .....CTGTTGGGTATAGATGACCT 1195
100 uIleThr...ProAspAspSerThrValSerLysAsp..... 111
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1196 GGAACCATGCGCATGACATGCTTCTGACCACCTGGATGATGATGATC 1245
112 .....IleLeuSerSerThrLeuGlnPheProThrGlnProVal 124
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1246 TCTTTGCCCCCTAGTCCAGGAGACTAATAAGCAGCCGCCCCAGACGGTG 1295
125 AsnIleProLeuTyraIaSerHisGlyAlaGluAspPheSerAlaGluTh 141
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1296 AAC...CCAATT..... 1304
141 rGluPheGluAsnHisLeuSerProAspSer.....ProGluG 155
: : : : : : : : : : : : : : : : : : : : : : : : : :
1305 .....GCCATCTC.....CCAGAAAGTTTAAACAAACCCGACC 1338
155 InValAlaProValIleAsnLeuGluProValGluLeuThrAlaSerHis 171
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1339 AGGTGGCCCCCTTCACTTCTTACAACTCTTCCCTTCCCTCCCGGGGTC 1388
172 MetThrValIleSerProAspGlyLeuLeuGlyMetGluLeuAlaSe 188
: : : : : : : : : : : : : : : : : : : : : : : : : :
1389 CTGTCC...TCCACTCCAGATCATTCCTTT...AGTTTAGAGCTGGGGC 1432
188 rGluSerLeuThrPheThrGluLeuAspPheValAsnPheAsnAspSera 205
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1433 TGAA...GTGGATATCACTGAAGAGATAGGAAGCCAGCTACACTGCTT 1479
205 laValGlySerIleGlyAlaGluLeuLeuGlySerProLeuSer 221
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1480 ACGTTGCCATGATC.....CCTCAGTGC 1502

```

```

222 ValAspAspValGluSerThrIleSerPheSerGlyProSerSerProGl 238
: : : : : : : : : : : : : : : : : : : : : : : : : :
1503 ATAAAGGAGGAAGAC.....ACCCCTTC 1525
238 uThrSerGlnSerSerIleIleGluSerSerProGluLeuTyrlsValI 255
: : : : : : : : : : : : : : : : : : : : : : : : : :
1526 AGATAATGATAGTGGCATCTGT...ATGAGCCAGAGTCCTATCTGTG... 1568
255 leSerThrSerSerIleAspAlaSerLysArgPheSerProTySerArg 271
: : : : : : : : : : : : : : : : : : : : : : : : : :
1569 .....GGGTCTCTCTCAGCACAGCCCTCTACCAGG 1598
272 SerSerLysSerLysGlnSerValIleThrSerAspAlaLysAlaProAr 288
: : : : : : : : : : : : : : : : : : : : : : : : : :
1599 GGTCTCCCAATAGG...AGCTCCCATCTCCAGGTGTCTCTGTGGGTC 1645
288 gLysThrArgThrProAlaGlnProValProGluHisValIleMet.... 303
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1646 TCCCGTCCCAACCTTACGATCTCTCTGGAGAGAAGATGTCAGCA 1695
304 .....GluHisLeuAspLysLysAspArgLysLysLeuGlnAsn 316
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1696 AAGTAAGGGTGAGAACTGGATGAAGAGTGAAAAAATGGAGCAAAAC 1745
317 LysAsnAlaAlaIleArgTyArgMetLysLysLysGlyGluAlaGlnGl 333
: : : : : : : : : : : : : : : : : : : : : : : : : :
1746 AAGACAGCAGCCACTAGTACCGCCAGAGAAGAGGGCGGAGCAGGAGGC 1795
333 yIleLysGlyGluGluGlnGluLeuGluGluLeuAsnThrLysLeuLysT 350
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1796 TCTTACTGTGAGTGCACAAAGAGCTGGAAGAAGAACGAGGCTCTAAAAG 1845
350 hrLysValAspAspLeuGlnArgGluIleLysTyMetLysAsnLeuMet 366
: : : : : : : : : : : : : : : : : : : : : : : : : :
1846 AGAGGGCGGATTCCTCCCAAGGAGATCCAGTACCTCGAAAGATTGATA 1895
367 GluAspValCysLysAlaLysGly 374
: : : : : : : : : : : : : : : : : : : : : : : : : :
1896 GAAGAGTCCCGCAAGGCAAGGGGG 1919

```

seq_name: gb_htg19:AC015801

seq_documentation_block:

LOCUS AC015801 176297 bp DNA HTG 26-MAR-2000
DEFINITION Homo sapiens chromosome 17 clone RP11-854A13 map 17, *** SEQUENCING
IN PROGRESS ***, 26 unordered pieces.

ACCESSION AC015801

VERSION AC015801.3 GI:7329323

KEYWORDS HTG; HTGS_PHASE1.

SOURCE human;

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 176297)

AUTHORS

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE

Homo sapiens chromosome 17, clone RP11-854A13

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 176297)

AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Collangelo,M., Collins,S., Collymore,A.,
Cooper,P., DeAtrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Hagan,B.,
Hagopian,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karat,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,K., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirrell,A., Vassiliev,H., Vo.A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 26, 2000 this sequence version replaced gi:6563606.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2128
Center clone name: 854_A_13

*** NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 48: contig of 48 bp in length
* 49 148: gap of 100 bp
* 149 1183: contig of 1035 bp in length
* 1184 1283: gap of 100 bp
* 1284 3141: contig of 1858 bp in length
* 3142 3241: gap of 100 bp
* 3242 4983: contig of 1742 bp in length
* 4984 5083: gap of 100 bp
* 5084 7035: contig of 1952 bp in length
* 7036 7135: gap of 100 bp
* 7136 10589: contig of 3454 bp in length
* 10590 10689: gap of 100 bp
* 10690 14230: contig of 3541 bp in length
* 14231 14330: gap of 100 bp
* 14331 17798: contig of 3468 bp in length
* 17799 17898: gap of 100 bp
* 17899 20725: contig of 2827 bp in length
* 20726 20825: gap of 100 bp
* 20826 24122: contig of 3297 bp in length
* 24123 24222: gap of 100 bp
* 24223 27097: contig of 2875 bp in length
* 27098 27197: gap of 100 bp
* 27198 32876: contig of 5679 bp in length
* 32877 32976: gap of 100 bp
* 32977 39229: contig of 6253 bp in length
* 39230 39329: gap of 100 bp
* 39330 44846: contig of 5517 bp in length
* 44847 44946: gap of 100 bp
* 44947 51703: contig of 6757 bp in length
* 51704 51803: gap of 100 bp
* 51804 58441: contig of 6638 bp in length
* 58442 58541: gap of 100 bp
* 58542 65154: contig of 6613 bp in length
* 65155 65254: gap of 100 bp
* 65255 73330: contig of 8076 bp in length
* 73331 73430: gap of 100 bp
* 73431 82270: contig of 8840 bp in length
* 82271 82370: gap of 100 bp
* 82371 92695: contig of 10325 bp in length
* 92696 92795: gap of 100 bp
* 92796 101707: contig of 8912 bp in length
* 101708 101807: gap of 100 bp
* 101808 110885: contig of 9078 bp in length
* 110886 110985: gap of 100 bp
* 110986 123688: contig of 12703 bp in length
* 123689 123788: gap of 100 bp
* 123789 140525: contig of 16737 bp in length
* 140526 140625: gap of 100 bp
* 140626 159784: contig of 19159 bp in length

* 159785 159884: gap of 100 bp
* 159885 176297: contig of 16413 bp in length.
FEATURES
source
1. 176297
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-854A13"
/clone_lib="RPC1-11 Human Male BAC"
misc_feature
1. 48
clone_end:17
vector_side:right
149. 1183
/note="assembly_fragment"
1284. 3141
/note="assembly_fragment"
3242. 4983
/note="assembly_fragment"
5084. 7035
/note="assembly_fragment"
7136. 10589
/note="assembly_fragment"
10690. 14230
/note="assembly_fragment"
14331. 17798
/note="assembly_fragment"
17899. 20725
/note="assembly_fragment"
20826. 24122
/note="assembly_fragment"
24223. 27097
/note="assembly_fragment"
clone_end:SP6
vector_side:right
27198. 32876
/note="assembly_fragment"
32977. 39229
/note="assembly_fragment"
39330. 44846
/note="assembly_fragment"
44947. 51703
/note="assembly_fragment"
51804. 58441
/note="assembly_fragment"
58542. 65154
/note="assembly_fragment"
65255. 73330
/note="assembly_fragment"
73431. 82270
/note="assembly_fragment"
82371. 92695
/note="assembly_fragment"
92796. 101707
/note="assembly_fragment"
101808. 110885
/note="assembly_fragment"
110986. 123688
/note="assembly_fragment"
123789. 140525
/note="assembly_fragment"
140626. 159784
/note="assembly_fragment"
159885. 176297
/note="assembly_fragment"
BASE COUNT 44519 a 42534 c 43191 g 43536 t 2517 others
ORIGIN
alignment_scores:
Quality: 262.00 Length: 401
Ratio: 1.278 Gaps: 20

```

175400 T...ATGAGCCCGCAGAGTCCCTATCTCTG. ....G 175378

262 laserLysArPheSerProTySerArGSerSerLysSerLysGlnSer 278
:|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
175377 GGTCTCCTCAGCATAGCCCTCTACGAGGGCTCTCCAAATAGG...AGC 175331

279 VallysThrSerAspAlaLysAlaProArgLysThrArgThrProAlaGln 295
:|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
175330 CTCCCATCTCCAGGTGTTCTCTGTGGGTCTGCCACCACCCAAACCTTACGA 175281

295 nProValProGluHisValIleMet. ....GluHisLeuA 307
:|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
175280 TCCCTCTGGAGAGAGATGGTAGCAGCAAAAGTAAAGGGTGAAGAACTGG 175231

307 spLysLysAspArgLysLysLeuGlnAsnLysAsnAlaAlaIleArgTyr 323
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
175230 ATAGAGCTCAAAAATAATGGAGCAAAACAGACAGCAGCCACTAGGTAC 175181

324 ArgMetLysLysGlyGluAlaGlnGlyIleLysGlyGluGlnGlnG 340
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
175180 CGCCAGAGAAGAGGGCGGAGCAGGAGGCTCTCACTGGCGAGTGCAAGA 175131

340 uLeuGluLeuLeuAsnThrLysLeuLysThrLysValAspLeuGlnA 357
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
175130 GCTGGAAAAGAACGACGAGGCTCTAAAAGAGAGGGCAGATCCCTGGCCA 175081

357 rGgGluIleLysTyrMetLysAsnLeuMetGluAspValCysLysAlaLys 373
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
175080 AGGAGATCCAGTACTGTAAGATTTGATAGAGAGGTCGCGAAGGCAAGG 175031

374 GAY 374
|||

175030 GGG 175028

seq_name: gb_htg30:AC021196

seq_documentation_block:
LOCUS AC021196 209861 bp DNA HTG 21-APR-2000
DEFINITION Homo sapiens chromosome 17 clone RP11-794C22, WORKING. DRAFT
SEQUENCE, 31 unordered pieces.
ACCESSION AC021196
VERSION AC021196.3 GI:7631117
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
2 (bases 1 to 209861)
Waterston,R.H.
Direct Submission
Submitted (14-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Apr 21, 2000 this sequence version replaced gi:7577693.

COMMENT
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0794C22
----- Summary Statistics -----
Sequencing vector: M13; 91%
Sequencing vector: plasmid; 9%
Chemistry: Dye-primer Et; 91% of reads
Chemistry: Dye-terminator Big Dye; 9% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 188997 bases at least Q40
Consensus quality: 195132 bases at least Q30
Consensus quality: 198570 bases at least Q20

```


Insert size: 225000; agarose-fp
 Insert size: 206861; sum-of-contigs
 Quality coverage: 3.50 in Q20 bases; agarose-fp
 Quality coverage: 3.75 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 31 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

* 1 2315: contig of 2315 bp in length
* 2316 2415: gap of unknown length
* 2416 4572: contig of 2157 bp in length
* 4573 4672: gap of unknown length
* 4673 7009: contig of 2337 bp in length
* 7010 7109: gap of unknown length
* 7110 9765: contig of 2656 bp in length
* 9766 9865: gap of unknown length
* 9866 12473: contig of 2608 bp in length
* 12474 12573: gap of unknown length
* 12574 16666: contig of 4093 bp in length
* 16667 16766: gap of unknown length
* 16767 20013: contig of 3247 bp in length
* 20014 20113: gap of unknown length
* 20114 24008: contig of 3895 bp in length
* 24009 24108: gap of unknown length
* 24109 28049: contig of 3941 bp in length
* 28050 28149: gap of unknown length
* 28150 31940: contig of 3791 bp in length
* 31941 32040: gap of unknown length
* 32041 35923: contig of 3883 bp in length
* 35924 36023: gap of unknown length
* 36024 40326: contig of 4303 bp in length
* 40327 40426: gap of unknown length
* 40427 45270: contig of 4844 bp in length
* 45271 45370: gap of unknown length
* 45371 48815: contig of 3445 bp in length
* 48816 48915: gap of unknown length
* 48916 53904: contig of 4989 bp in length
* 53905 54004: gap of unknown length
* 54005 58337: contig of 4333 bp in length
* 58338 58437: gap of unknown length
* 58438 63057: contig of 4620 bp in length
* 63058 63157: gap of unknown length
* 63158 67683: contig of 4526 bp in length
* 67684 67783: gap of unknown length
* 67784 74319: contig of 6536 bp in length
* 74320 74419: gap of unknown length
* 74420 83151: contig of 8732 bp in length
* 83152 83251: gap of unknown length
* 83252 89129: contig of 5878 bp in length
* 89130 89229: gap of unknown length
* 89230 98220: contig of 8991 bp in length
* 98221 98320: gap of unknown length
* 98321 105559: contig of 7239 bp in length
* 105560 105659: gap of unknown length
* 105660 113081: contig of 7422 bp in length
* 113082 113181: gap of unknown length
* 113182 120975: contig of 7794 bp in length
* 120976 121075: gap of unknown length
* 121076 132903: contig of 11828 bp in length
* 132904 133003: gap of unknown length
* 133004 144491: contig of 11488 bp in length
* 144492 144591: gap of unknown length
* 144592 153726: contig of 9135 bp in length
* 153727 153826: gap of unknown length
* 153827 171330: contig of 17504 bp in length
* 171331 171430: gap of unknown length
* 171431 191919: contig of 20489 bp in length
* 191920 192019: gap of unknown length

```

```

FEATURES             * 192020 209861: contig of 17842 bp in length.
    source            Location/Qualifiers
                        1..209861
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="17"
                        /clone="RP11-794C22"
BASE COUNT      52775 a 51288 c 51341 g 51413 t 3044 others
ORIGIN
alignment_scores:
    Quality: 262.00      Length: 401
    Ratio: 1.278        Gaps: 20
    Percent Similarity: 51.122 Percent Identity: 28.180
alignment_block:
    US-08-656-811A-1 x AC021196 ..
Align seg 1/1 to: AC021196 from: 1 to: 209861
19 GluMetProValValGlnThrAspGlyGlnPheGlyAspLeuLysSer... 34
||||| : : : : : : : : : : : : : : : : : : : : : : : :
196107 GAATGAGCTTCCTGAGCAACGAGGTGTGGTGGGGGACTTGATGTC 196156
35 .....ThrSerArgHisGlyGlyAspGluSerLeuSerLeu..... 46
||||| : : : : : : : : : : : : : : : : : : : : : : : :
196157 CTTCCAGCAGTCGGGTGTTGGGGCTGAAGAAAGCATAGTCTCTT 196206
47 .....: : : : : : : : : : : : : : : : : : : : : : : :
196207 ACTAGTGGAGTGCCCAAGCACCTTCAACACCTCATGGG..... 196244
55 LeuGluProPheGluGluAspValLeuGlyAla.....GluTrpMe 68
||||| : : : : : : : : : : : : : : : : : : : : : : : :
196245 .....TTCCTCAGCGACAAAGGCTAAGCGGGGCTCTCCGATGGCT 196285
68 tGluSerSerAspLeu.....GlySerPheLeuAspAlaL 80
: : : : : : : : : : : : : : : : : : : : : : : :
196286 GACTGTGGATGGTGGTTCAGTCCCTCCAAACAACAGAGGAGGATGCT 196335
80 euGlyAspAsnHisGluArgLeuHisProPheGluSerAsnLeuGlu 96
: : : : : : : : : : : : : : : : : : : : : : : :
196336 TCTCTGGGACACATGGATGTGGAGAAATGGATTTGAAGGAGTTCGAC 196385
97 PheThrSerLeuLeuThrProAspAsp...SerThrValSerLysAsp 112
||| : : : : : : : : : : : : : : : : : : : : : : : :
196386 TTTGGTGCCTGTTGGGTATAGATGACCTGGAAACCATGCCAGTGACCT 196435
112 eLeuSerSer.....: : : : : : : : : : : : : : : : : :
196436 TTTGACCACGTTGGATGACACTTGTAAATCTCTTTGCCCCCTTAGTCCAGG 196485
116 ..ThrLeuGlnPheProThrGlnProValAsnIleProLeuThrAlaSer 131
||| : : : : : : : : : : : : : : : : : : : : : : : :
196486 AGACTAATAAGGTGCCCCCGCAGATGGTGAAC...CCAATT..... 196523
132 HisGlyAlaGluAspPheSerAlaGluThrGluPheGluAsnHisLeu 148
: : : : : : : : : : : : : : : : : : : : : : : :
196524 .....: : : : : : : : : : : : : : : : : : : : : : : :
148 rProProAspSer.....ProGluGlnValAlaProValIleAsnL 162
||||| : : : : : : : : : : : : : : : : : : : : : : : :
196533 ....CCAGAAAGTTTAAACAAACCCGACGAGTTCGCCCTTCACCTTCT 196578
162 euGluProValGluLeuThrAlaSerHisMetThrValIleSerProAsp 178
: : : : : : : : : : : : : : : : : : : : : : : :
196579 TGCACACTTCTTCCCTCCCGCAGGG...GTCCAGTCTCTCCACATCCAGAT 196625
179 GlyLeuLeuGlyGlyMetGluLeuAlaSerGluSerLeuThrPheThrCl 195
: : : : : : : : : : : : : : : : : : : : : : : :
196626 CATTCCTTT...AGTTTAGAGCTGGGCACTGAA...GTGGATATCACTGA 196669

```

```

195 uLeuAspPheValAsnPheAsnAspSerAlaValGlySerIleGlyGlyA 212
196670 AGAAGATAGGAAGCGGACTCCACTGCTTACGTTGCCATGATC..... 196712
212 laGluGluLeuGlySerProLeuSerValAspValGlySerThr 228
196713 .....CCTCAGTCATAAAGGAGGAGAAC..... 196736
229 IleSerPheSerGlyProSerSerProGluThrSerGlnSerIleI 245
196737 .....ACCCCTCAGATAATGATGTCGATCTG 196765
245 eGluSerSerProGluLeuTyrlsValIleSerThrSerSerIleAspA 262
196766 T...ATGAGCCACAGTCCTATCTG.....G 196788
262 laSerLysArgPheSerProTyrlsSerArgSerSerLysSerLysGlnSer 278
196789 GGTCTCTCAGCATAGCCCTCTACCAAGGGCTCTCCAAATAGG...AGC 196835
279 ValLysThrSerAspAlaLysAlaProArgLysThrArgThrProAlaG 295
196836 CTCCACTCTCCAGTGTCTCTGTGGTCTGCCACCACCAACCTTACGA 196885
295 nProValProGluHisValIleMet.....GluHisLeuA 307
196886 TCCTCTCGGAGAGAGAGTGGTAGCAGCAAAAGTAAGGGTGAGAACTGG 196935
307 spLysLysAspArgLysLysLeuGlnAsnLysAsnAlaIleArgTyr 323
196936 ATAAGAAGCTGAAAAAATGAGCAAAACAGACAGCAGCCACTAGGTAC 196985
324 ArgMetLysLysLysGlyGluAlaGlnGlyIleLysGlyGluGluGln 340
196986 CCCAGAGAGAGAGGGGAGCAGCAGAGGCTCTCAGTCGCGAGTGCARA 197035
340 uLeuGluGluLeuAsnThrLysLysLeuLysThrLysValAspAspLeuGlnA 357
197036 GCTGAAAAGAGAGAGAGGCTCTAAAAGAGAGGGCAGATCCCTGGCCA 197085
357 rgGluIleLysTyrMetLysAsnLeuMetGluAspValCysLysAlaLys 373
197086 AGGAGATCCAGTACCTGAAGATTGTAGAGAGGTCCTCCGAGGCAAGG 197135
374 Gly 374
197136 GGG 197138
seq_name: gb_pr2:HS1104E15
seq_documentation_block:
LOCUS HS1104E15 112460 bp DNA PRI 12-DEC-1999
DEFINITION Human DNA sequence from clone RP5-1104E15 on chromosome
22q12.3-13.1. Contains the MGAT3 gene for mannosiyl
(beta-1,4)-glycoprotein beta-1,4-N-acetylglucosaminyltransferase,
the gene for a predicted protein, the ATE4 gene for activating
transcription factor 4 (tax-responsive enhancer element B67) and
the 5' end of the CACNA1I gene for voltage-dependent calcium
channel, alpha 1I subunit. Contains ESTs, STSS, GSSs and five
putative CpG islands, complete sequence.
ACCESSION AL022312
VERSION AL022312.7 GI:4914501
KEYWORDS HTG; ATE4; CACNA1I; calcium channel; CpG island; MGAT3.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 112460)
AUTHORS Barlow,K.
TITLE Direct Submission
JOURNAL Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk

```

On May 28, 1999 this sequence version replaced gi:490589. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep RP5-1104E15 is from the library RPCI-5 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>

VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone RP5-1104E15. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone RP5-1104E15 is at 1 in this sequence. The true left end of clone CTA-206C7 is at 112356 in this sequence. The true right end of clone RP3-407F17 is at 83950 in this sequence. The end of this sequence overlaps with sequence AL008716.

FEATURES

Location/Qualifiers

1..112460

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosomes="22"

/map="q12.3-13.1"

/clone="RP5-1104E15"

/clone_lib="RPCI-5"

1..62

/note="Alu repeat: matches 1..62 of consensus"

82..149

/note="MIR repeat: matches 75..140 of consensus"

150..440

/note="AluJo repeat: matches 17..303 of consensus"

441..556

/note="MIR repeat: matches 140..262 of consensus"

1072..1198

/note="MIR repeat: matches 13..143 of consensus"

1205..1511

/note="AluX repeat: matches 1..308 of consensus"

1751..1841

/note="MER63 repeat: matches 26..752 of consensus"

1891..1931

/note="L2 repeat: matches 2664..2704 of consensus"

2297..2644

/note="match: GSS: Em:AQ218889"

2514..2569

/note="MIR repeat: matches 113..170 of consensus"

2592..2887

/note="AluSc repeat: matches 1..298 of consensus"

3001..3085

/note="L2 repeat: matches 2619..2710 of consensus"

3088..3127

/note="5 copies 8 mer atgaatga 92 conserved"

3839..4037

/note="MIR repeat: matches 31..248 of consensus"

4197..4326

/note="MER5A repeat: matches 20..149 of consensus"

4343..4626

```

repeat_region /note="AluJo repeat: matches 25. .305 of consensus"
5079. .5125
/note="L2 repeat: matches 2686. .2735 of consensus"
repeat_region 5126. .5429
/note="AluSx repeat: matches 1. .303 of consensus"
repeat_region 5439. .5743
/note="AluY repeat: matches 1. .300 of consensus"
repeat_region 6365. .6649
/note="AluSx repeat: matches 27. .311 of consensus"
repeat_region 7636. .7708
/note="L2 repeat: matches 2638. .2710 of consensus"
repeat_region 8325. .8386
/note="31 copies 2 mer tt 72 conserved"
repeat_region 8328. .8387
/note="L2 copies 5 mer tcttt 86 conserved"
repeat_region 8412. .8493
/note="Alu repeat: matches 225. .308 of consensus"
repeat_region 9387. .9483
/note="L2 repeat: matches 2611. .2710 of consensus"
repeat_region 10720. .11321
/note="L2 repeat: matches 2065. .2709 of consensus"
repeat_region 11479. .11599
/note="MIR repeat: matches 31. .145 of consensus"
repeat_region 12182. .12267
/note="43 copies 2 mer cc 64 conserved"
repeat_region 13334. .13419
/note="MIR repeat: matches 44. .144 of consensus"
repeat_region 13871. .14283
/note="MSTD repeat: matches 1. .391 of consensus"
misc_feature complement(15277. .15778)
repeat_region 13818. .15839
/note="11 copies 2 mer tt 100 conserved"
misc_feature complement(15839. .16591)
misc_feature /note="match: GSS: Em:AQ740632"
repeat_region 16460. .16520
/note="match: GSS: Em:AQ740620"
misc_feature complement(16461. .17266)
repeat_region 16521. .16647
/note="match: GSS: Em:AQ739191"
repeat_region 17352. .17478
/note="FLAN_A repeat: matches 1. .133 of consensus"
repeat_region 18248. .18455
/note="MIR repeat: matches 32. .148 of consensus"
repeat_region 18484. .18655
/note="MIR repeat: matches 51. .262 of consensus"
repeat_region 19753. .20050
/note="MIR repeat: matches 84. .256 of consensus"
repeat_region 20336. .20350
/note="AluJo repeat: matches 1. .293 of consensus"
repeat_region 20351. .20410
/note="MIR repeat: matches 246. .259 of consensus"
repeat_region 20411. .20476
/note="L2 repeat: matches 2666. .2729 of consensus"
repeat_region 20554. .20642
/note="MIR repeat: matches 40. .246 of consensus"
repeat_region 20687. .20747
/note="L2 repeat: matches 2531. .2620 of consensus"
repeat_region 21468. .22247
/note="MIR repeat: matches 93. .153 of consensus"
misc_feature complement(22233. .22544)
repeat_region 22466. .22617
/note="match: GSS: Em:AQ594604"
repeat_region 22640. .22977
/note="MIR repeat: matches 82. .262 of consensus"
repeat_region 23096. .23274
/note="AluJb repeat: matches 1. .300 of consensus"
repeat_region 23902. .24207
/note="MIR repeat: matches 10. .192 of consensus"
/note="AluSp repeat: matches 1. .312 of consensus"

```

```

repeat_region 24743. .24836
/note="MIR repeat: matches 35. .142 of consensus"
repeat_region 24996. .25049
/note="MIR repeat: matches 92. .148 of consensus"
repeat_region 25050. .25361
/note="AluSx repeat: matches 1. .312 of consensus"
repeat_region 25410. .25700
/note="AluSx repeat: matches 1. .291 of consensus"
repeat_region 26006. .26197
/note="L1MC3 repeat: matches 7546. .7734 of consensus"
repeat_region 26198. .26512
/note="AluSx repeat: matches 1. .312 of consensus"
mRNA 26520. .32881
/gene="MGAT3"
/note="match: cDNAs: Em:E11285; match: ESTs: Em:AI452685
Em:AI524129 Em:AI198445 Em:AI439025 Em:AI075278
Em:AA743691 Em:AA421504 Em:AI090769 Em:AI308859
Em:AA829826 Em:AI571173 Em:AA501937 Em:AA421473
Em:AA326081 Em:H19142 Em:H29553 Em:AA345131 Em:H47026
Em:AI572262 Em:AA876086 Em:AI364862 Em:AI367012 Em:H92616
Em:H41010"
/evidence="not_experimental"
/product="dJ1104E15.1 (mannosyl (beta-1.4-)-glycoprotein
beta-1,4-N-acetylglucosaminyltransferase)"
gene 26520. .32881
/gene="MGAT3"
repeat_region 26834. .27125
/note="AluSg repeat: matches 7. .298 of consensus"
repeat_region 27188. .27489
/note="AluJb repeat: matches 1. .302 of consensus"
misc_feature 28035. .30123
/gene="MGAT3"

alignment_scores:
Quality: 261.00 Length: 340
Ratio: 1.442 Caps: 19
Percent Similarity: 53.235 Percent Identity: 30.882

alignment_block:
US-08-656-811A-1 x HS1104E15 ..

Align seg 1/1 to: HS1104E15 from: 1 to: 112460

55 LeuGluProPheGluGluAspValLeu...GlyAlaGluTrpMet...G1 69
||||| ||| |||||: |||: |||: |||: |||: |||: |||
62446 CTCAAATGTTTGCAGAGGATGCTCTCCGGGACAGATGGATCTGGA 62495
||: ||||| :||| |||||
62496 GAAATGGATTGAAGAGGATTCGACTTGGATGCC..... 62529
||||| :||| :||| :||| :||| :||| :||| :|||
85 luArgLeuHisProPheGluSerAsnLeuLeuGluPheThrSerLeuIle 101
||||| :||| :||| :||| :||| :||| :||| :|||
62530 .....CTGTTGGGTATAGTACCTGGAA 62553

102 Thr...ProAspSerThrValSerLysAsp..... 111
||| ||||| |||: |||: |||: |||: |||: |||: |||
62554 ACCATGCCAGATGACCTTCTGACCACGTTGGATGACACTTCTGATCTCTT 62603
||||| |||: |||: |||: |||: |||: |||: |||
112 .....lleLeuSerSerThrLeuGlnPheProThrGlnProValAsnI 126
|||||: |||: |||: |||: |||: |||: |||: |||
62604 TGCCCCCTAGTCCAGGAGACTAATAAGCAGCCCCCAGCGGTGAAC. 62652
||||| |||: |||: |||: |||: |||: |||: |||: |||
126 leProLeuTyAlaSerHisGlyAlaGluAspPheSerAlaGluThrGlu 142
||||| :|||: |||: |||: |||: |||: |||: |||
62653 ..CCAATT..... 62658

143 PheGluAsnHisLeuSerProProAspSer.....ProGluGlnVa 156
||||| |||: |||: |||: |||: |||: |||: |||: |||
62659 .....GGCCATCTC.....CCAGAAAGTTTAAACAAACCCGACGAGT 62696
||||| |||: |||: |||: |||: |||: |||: |||: |||
156 lAlaProValIleAsnLeuGluProValGluLeuThrAlaSerHisMet 173
||||| |||: |||: |||: |||: |||: |||: |||: |||

```

```

62697 TGCCCCCTTACCTCTTCTTACAACTCTTCCCTTTCCCGGGGCTCTGT 62746
173 hrValIleSerProAspGlyLeuLeuGlyGlyMetGluLeuAlaSerGlu 189
   :::::::::::::::::::::
62747 CC...TCCACTCCAGATCATCTCTT...AGTTAGACTGGCAGTGAA 62790
190 SerLeuThrPheThrGluLeuAspPheValAsnPheAsnAspSerAlaVa 206
   :::::::::::::::::::::
62791 ...GTGGATATCACTGAAGGAGATAGGAAGCAGACTACACTGCTTACGT 62837
206 lGlySerIleGlyGlyAlaGluLeuLeuGlySerProLeuSerVala 223
   :::::::::::::::::::::
62838 TGCCATGATC.....CCTCAGTGCATAA 62860
223 sPaspValIGlySerThrIleSerPheSerGlyProSerProGluThr 239
   :::::::::::::::::::::
62861 AGGAGGAAGAC.....ACCCCTCAGAT 62883
240 SerGlnSerSerIleIleGluSerSerProGluLeuThrLysValIleSe 256
   :::::::::::::::::::::
62884 AATCATAGTGGCACTGT...ATGAGCCACAGAGCTCTATCTG..... 62922
256 rThrSerSerIleAspAlaSerLysArgPheSerProTyrSerArgSerS 273
   :::::::::::::::::::::
62923 .....GGGTCTCTCCAGACAGCCCTCTTACCAGGGCT 62956
273 erLysSerLysGlnSerValLysThrSerAspAlaLysAlaProArgLys 289
   :::::::::::::::::::::
62957 CTCCAATPAG...AGCTCCATCTCCAGGTGTCTCTGGGTCTGCC 63003
290 ThrArgThrProAlaGlnProValProGluHisValIleMet..... 303
   :::::::::::::::::::::
63004 CGTCCCAACCTTACGATCTCTCTGAGAGAAGATGGTAGCAGCAAGAGT 63053
304 .....GluHisLeuAspLysLysAspArgLysLysLeuGlnAsnLysA 318
   :::::::::::::::::::::
63054 AAAGGGTGAGAACTGGATAGAAGCTGAAAAAATGGAGCAAAACAGA 63103
318 snAlaAlaIleArgTyrArgMetLysLysGlyGluAlaGlnGlyIle 334
   :::::::::::::::::::::
63104 CAGCAGCCACTAGTACCGCAGAGAAGAGCGCGGAGGAGGCTCTT 63153
335 LysGlyGluGluGlnGluLeuGluLeuAsnThrLysLeuLysThrLys 351
   :::::::::::::::::::::
63154 ACTGGTGAGTCAAAAGAGCTGGAAGAAAGAACGAGGCTCTAAAGAGAG 63203
351 sValAspAspLeuGlnArgGluIleLysTyrMetLysAsnLeuMetGluA 368
   :::::::::::::::::::::
63204 GCGGATTCCTCGCCAGGAGATCCAGTACCTGAAAGATTGATAGAAG 63253
368 spValCysLysAlaLysGly 374
   :::::::::::::::::::::
63254 AGTCCCGAAGCAAGGGG 63273

```

seq_name: gb_ro:MUSTRANFAC

```

seq documentation_block:
LOCUS MUSTRANFAC 1224 bp mRNA ROD 23-JUL-1993
DEFINITION Mus musculus (clone Chopil) C/ATF mRNA, complete cds.
ACCESSION L13791
VERSION L13791.1 GI:293841
KEYWORDS C/ATF; DNA-binding protein; transcription factor.
SOURCE Mus musculus (library: lambda-ZAP of B. Spiegelman) cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1224)
AUTHORS Vallejo,M., Ron,D., Miller,C.P. and Habener,J.F.
TITLE C/ATF, a member of the activating transcription factor family of
DNA-binding proteins, dimerizes with CAAAT/enhancer-binding proteins
and directs their binding to cAMP response elements
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 90, 4679-4683 (1993)
MEDLINE 93281642

```

FEATURES

```

source
1..1224
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="NIH-3T3-L1"
/cell_type="adipocyte"
/tissue_lib="lambda-ZAP of B. Spiegelman"
100..1149
/standard_name="C/ATF"
/codon_start=1
/function="transcription factor"
/evidence=experimental
/product="C/ATF"
/protein_id="AAA40476.1"
/db_xref="GI:293842"
/translation="MTMSFLNSEVLGDLMSPFDSGLGABESLGLLDDYLEVAKHL
KHPFSSDRAGSEWAMDDGLASDSTGKDFASGTDMLEKMDLKEFDLALFRMD
DLETMPDELLTLDLTDCLFAPLVQETNKEPQTNPVNIHGLPESLKVQVAPFTFLQ
PPCPSPGLSSTPSHSELSLEAGVDISEGDRKPSAAVITLIPCKVEDPDSNDS
GICMSPESYLGSPHSPTSRAPPDNLSPGSGSRGSPKPYDPGVSILTAKVKTEKL
DKLKKMEONKTAATRYRQKRAEQALTGECKELEKKNKALKEKADSLAKEIQYLKD
LIEVRKARGQKRP"
polyA_signal
1206..1211
/note="putative"
BASE COUNT 321 a 294 c 326 g 283 t
ORIGIN

alignment_scores
Quality: 259.50 Length: 433
Ratio: 1.324 Gaps: 19
Percent Similarity: 45.266 Percent Identity: 27.252

alignment_block:
US-08-656-811A-1 x MUSTRANFAC
Align seg 1/1 to: MUSTRANFAC from: 1 to: 1224

13 AlaArgGluThrPGlyLeuGluMetProValValGlnThrAspGlyGlnPh 29
|||
|||
|||
67 GCACATCTCTGGATCCAGCAAGCCCAACATGACCGAGATGAGCTT 116
29 e.....GlyAspLeuLysSer.....Thr 36
|||
117 CCTGAACAGCAAGCTGTGGCGGGGACTTGTATGTCCTCCCTTCGACCACT 166
36 erArgHisGlyGlyAspGluSerLeuSerLeuGlnProGln..... 49
|||
167 CGGCTTTGGGGCTGAGAAAGCCTAGTCTCTTAGATGACTATCTGGAG 216
50 ...GlyAlaThrLeuLysLeuGluProPheGluGluAspValLeuGly... 64
|||
217 GTGCCCAAGCACTTGAACCTCATGGGTCTCTCCAGCGCAAGCGGGCTC 266
65 .AlaGluTirp.....MetGluSerSerAspLeuG 74
|||
267 CTCGGAATGCCGCTATGGATGATGGCTTGGCCAGTGCCTCAGACACCG 316
74 lySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeuHisProPhe 90
|||
317 GCAAG...GAGGATGCCTTTCCGGGAGAGATTTGGATGTTGGAGAAATG 363
91 GluSerAsnLeuLeuGluPheThrSerLeuIleThrProAspAspSerTh 107
|||
364 GATCTGAAGAGATTGTGACTTCGATGCTCTGTTTCCGAATGGATGACCTGA 413
107 rValSerLysAspIleLeuSerSerThrLeu..... 117
|||
414 AACCATCCAGATGAGTCTTGACCACTGTTGGATGACACATGCTGATCTTT 463
118 .....GlnPheProThrGlnProValAsn 125
|||
464 TTGCCCTCTAGTCCAAGAGACTAATAAGGAGCCCTCAGACAGTGAAC 513

```

```
126 IleProLeuTyrAlaSerHisGlyAlaGluAspPheSerAlaGluThrCl 142
||||:
514 ...CCAATT..... 519
142 uPheGluAsnHisLeuSerProProAspSer.....ProGluGlnV 156
||||:
520 .....GCCCATCTC.....CGAAGAAGTTTATAAAAGTCGACGAG 556
156 aAlaProValIleAsnLeuGluPro..... 164
||||:
557 TTGCCCCCTTACATTCTGCAGCCTTCCCTGTTCCCGAGGTTCTG 606
165 .....ValGluLeu..... 167
607 TCTTCCACTCCAGACATTCTTTAGTTAGCTAGGCGAGTGAAGTTGA 656
168 .....ThrAlaSerHisMetThrValI 175
657 TATCTCTGACGAGACAGCAAGCCTGCTCTGCTTACATTCTCTAA 706
175 leSerProAspGlyLeuLeuGlyMetGluLeuAlaSerGluSerLeu 191
|||
707 TTCTCTCCA..... 714
192 ThrPheThrGluLeuAspPheValAsnPheAsnAspSerAlaValGlySe 208
|||
715 TGTGTAAGAGGAGAGACACTCTCTGCTGACATGACAGT.....GGCAT 758
208 rIleGlyGlyAlaGluLeuLeuGlySerProLeuSerValAspAspV 225
|||
759 CTGTATGACGGCGAGTCTTACCTGGGCTCTCC..... 792
225 aGluSerThrIleSerPheSerGlyProSerSerProGluThrSerGln 241
||||:
793 .....CACCATAGCCCTCCACCTCCAGG 816
242 SerSerIleIleGluSerSerProGluLeuTyrLysValIleSerThrSe 258
|||
817 .....GCCCACACAGAC.....AATCTGCCTTCTCCAGG 845
258 rSerIleAspAlaSerLysArgPheSerProTyrSerArgSerSerLysS 275
|||
846 TGGTTCCTGGGTCTCTCGGCCCAACCTTATGAC..... 882
275 erLysGlnSerValLysThrSerAspAlaLysAlaProArgLysThrArg 291
882 ..... 882
292 ThrProAlaGlnProValProGluHisValIleMetGluHisLeuAspLy 308
|||
883 CCACCTGGAGTTAGTTTGACAGCTAAAGTGAAGACTGAGAAATGGATAA 932
308 sLysAspArgLysLysLeuGlnAsnLysAsnAlaAlaIleArgTyrArgM 325
||||:
933 GAAGCTGAAGAAGATGAGACAAAAACAGACAGCAGCCACTAGGTACCCGC 982
325 etLysLysLysGlyGluAlaGlnGlyIleLysGlyGluGlnGluLeu 341
|||
983 AGAACAAGCGGCTGAGCAGAGGCGCCCTCACTGGCGAGTGAAGAGCTA 1032
342 GluGluLeuAsnThrLysLeuLysThrLysValAspAspLeuGlnArgGl 358
|||
1033 GAAAAAAGATGAGGCTCTGAAAGAAGAAGGAGATCTCTGCGCCAAGA 1082
358 uIleLysTyrMetLysAsnLeuMetGluAspValCysLysAlaLysGly 374
||||:
1083 GATCCAGTATCTGAAGACCTGATAGAAGGTCCGTAAGCGAAGGGG 1131
seq_name: gb_ro:MUSM4TF4
seq_documentation_block:
LOCUS MUSM4TF4 1391 bp mRNA ROD 09-NOV-1994
```

```
DEFINITION M.musculus mATF4 (mTR67) mRNA, complete cds.
ACCESSION M94087
VERSION M94087.1 GI:293723
KEYWORDS mATF4; mTR67.
SOURCE Mus musculus (strain CD-1) embryo cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Chevray,P.M. and Nathans,D.
TITLE Protein interaction cloning in yeast: Identification of mammalian
proteins that react with the leucine zipper of Jun
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (13), 5789-5793 (1992)
MEDLINE 92335183
COMMENT On Jun 12, 1993 this sequence version replaced gi:199822.
FEATURES
source
1..1391
/organism="Mus musculus"
/strain="CD-1"
/db_xref="taxon:10090"
/tissue_type="embryo"
271..1373
/gene="mATF4"
271..1311
/gene="mATF4"
/translation="MSFLNSEVLADGLMSPDQSLGAEBSGLDDLYLEVAKHLKPH
GFSSDKAGSEWPMDDGLASADFGSDWMLEKMDLKEFDLFRMDLLE
TMPDELLTLDLDDTCDLEAPLVQETNKPPQTNPIGHLPESLIKVDQVAPFTLQPPF
CSGVLSTTSEHSFSLGSEVDISEGDRKPDAAITLIPCVKEEDTPSDNDSGIC
MSPEYLGSQHSSTSRAPPDNLSPGSGSRSPRPYDPGPGVSLTAKYKTEKLDKK
LKMEQNKTAATRYRQRRAEQEALTGECELEKNEALKERAKDSLAKEIQYLKDLIE
EVRKARQKRRVP"
1368..1373
/gene="mATF4"
1391
/gene="mATF4"
BASE COUNT 343 a 354 c 377 g 317 t
ORIGIN
alignment_scores:
Quality: 259.50 Length: 433
Ratio: 1.324 Gaps: 19
Percent Similarity: 45.266 Percent Identity: 27.252
alignment_block:
US-08-656-811A-1 x MUSM4TF4 ..
Align seg 1/1 to: MUSM4TF4 from: 1 to: 1391
13 AlaArgGluTrpGlyLeuGluMetProValGlnThrAspGlyGlnPh 29
||||:||||:||||:||||:||||:||||:||||:||||:||||:
229 GCACATTCCTGGAATCCACAAAGCCCCACACATGACCGAGATGAGCTT 278
29 e.....GlyAspLeuLysSer.....ThrS 36
|
279 CCTGAACACGCAAGTGTGGCGGGGAGCTTGATGCCCTTCGACCACT 328
|
36 erArgHisGlyGlyAspGluSerLeuSerLeuGlnProGln..... 49
|
329 CGGGTTGGGGCTGAAGAAAGCTAGGTCTCTTAGATGACATATCTGGAG 378
50 ...GlyAlaThrLeuLysLeuGluProPheGluGluAspValLeuGly.. 64
|||
379 GTGGCCAAAGCACTTGAACCTCATGGGTTCTCCAGCGACAAGCGGCTC 428
65 .AlaGluTrp.....MetGluSerSerAspLeuG 74
||||:||||:||||:||||:||||:||||:||||:||||:||||:
|
```


BASE COUNT 334 a 354 c 368 g 309 t
ORIGIN

alignment_scores:
Quality: 255.00 Length: 434
Ratio: 1.288 Gaps: 21
Percent Similarity: 45.622 Percent Identity: 27.880

alignment_block:

US-08-656-811a-1 x MMATF4 ..

Align seg 1/1 to: MMATF4 from: 1 to: 1365

```

13 AlaArgGluTrpGlyLeuGluMetProValValGlnThrAspGlyGlnPh 29
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
240 GCACATTCTGGAATCCAGCAAGCCCAACACATGACCGAGATGAGCTT 289
29 e.....GlyAspLeu.....LysSerThrSerA 37
|.....|.....|.....|.....|.....|.....|.....|.....|
290 CCGTAACACGCGAAGTGTGGCGGGGACTTGATGACCCCTTCGACGAGTC 339
37 tq.HisGlyGly.AspGluSerLeuSerLeuGlnProGln..... 49
||.....|.....|.....|.....|.....|.....|.....|.....|
340 CGGTTGGGGCTGAAGAAAGCTAGGCTCTTAGATGACTATCTGGAGG 389
50 ..GlyAlaThrLeuLysLeuGluProPheGluGluAspValLeuGly... 64
:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
390 TGGCCAAGCAGCTGAACCTCTATGTTCTCCAGCAAGCGCGCTCC 439
65 AlaGluTrp.....MetGluSerSerAspLeuG1 74
:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
440 TCGGAATGGCGCGCTATGGATGANGCTTGGCCAGTGCCTCAGACACCGG 489
74 ySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeuHisProPheG 91
|:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
490 CAAG..GAGGATGCTTTTCGGGACAGATTGGATGTGGAGAAATGG 536
91 LuSerAsnLeuLeuGluPheThrSerLeuLeuThrProAspAspSerThr 107
||:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
537 ATCTCAAAAGAGTTTCACTTCGATGCTCTGTTTCGAATGGATGCACCTG 586
108 ValSerLysAspIleLeuSerSerThrLeu..... 117
:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
587 ACCATGCCAGATGAGCTCTTGACCCAGTTGGATGACACATGTGATCTTTT 636
118 .....GlnPheProThrGlnProValAsnI 126
637 TCCCCCTCTAGTCCAGAGACTAATAAGGAGCCCTCAGACAGTGAAC. 685
126 leProLeuTyraLaserHisGlyAlaGluAspPheSerAlaGluThrGlu 142
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
686 ..CAAATT..... 691
143 PheGluAsnHisLeuSerProProAspSer.....ProGluGlnVa 156
:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
692 .....GCCCATCTC.....CCAGAAAGTTTAATAAAGTCGACAGGT 729
156 lAlaProValIleAsnLeuPro..... 164
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
730 TGCCCCCTTACATTCTTGACGCTTCCCTGTTCCCGAGGGGTTCTGT 779
165 .....ValGluLeu..... 167
780 CTTCACCTCCAGAGCATTCCTTTAGTTTAGAGCTAGGCGATCAAGTTGAT 829
168 .....ThrAlaSerHisMetThrValI 175
:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
830 ATCTCTGAAGGAGACAGGAAGCCCTGACTCTGCTGCTTACATTACTTAAT 879
175 eSerProAspGlyLeuLeuGlyGlyMetGluLeuAlaSerGluSerLeuT 192
|.....|.....|.....|.....|.....|.....|.....|.....|
880 CCCTCCA.....T 887

```

```

192 hrPheThrGluLeuAspPheValAsnPheAsnAspSerAlaValGlySer 208
||| ||| ::| ||||| |||
888 GTGTAAAGGAGGAGACACTCCCTCTGTGACATGACAGT.....GCATC 931
209 IleGlyGlyAlaGluGluLeuGlySerProLeuSerValAspVa 225
::: ||||: ||||| |||||
932 TGTATGAGCGGAGTCTTACCTGGGCTCTCC..... 964
225 lGluSerThrIleSerPheSerGlyProSerSerProGluThrSerGlns 242
965 .....CAGCATAGCCCTCCACTCCAGG. 988
242 erSerIleGluSerSerProGluLeuTyrlLysValIleSerThrSer 258
::: ||||: ||||: ||| :::
989 .....GCCCCACCAGAC.....AATCTGCCTTCTCCAGGT 1018
259 SerIleAspAlaSerLysArgPheSerProTyrlSerArgSerLysSe 275
::: ||||: ||| ::||| ||||:
1019 GGTTCGCGTGGTCTCTCGGCCCAACCTTATGAC..... 1054
275 rLysGlnSerValLysThrSerAspAlaLysAlaProArgLysThrArgT 292
1055 .....C 1055
292 hrProAlaGlnProValProGluHisValIleMetGluHisLeuAspLys 308
|||::: ||| ||| ||||| |||
1056 CACCTGGAGTTAGTTTGACAGCTAAAGTCAAGACTCAGAAATGGATAAG 1105
309 LysAspArgLysLysLeuGlnAsnLysAsnAlaAlaIleArgTyrlArgMe 325
|||::: ||||| ||||| ||||| ||||| |||||
1106 AAGCTCAAAAGATGGAGCAACAAACAGACAGCCACTAGGTACCGCA 1155
325 tLysLysLysGlyGluAlaGlnGlyIleLysGlyGluGluGlnGluLeuG 342
:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1156 GAAGAGCGGGCTGAGCAGGAGGCCCTCACTGGCAGTGTAGGAGCTAG 1205
342 luGluLeuAsnThrLysLysLeuLysThrLysValAspAspLeuGlnArgGlu 358
|||::: ||| ||||| ||||:||||:||||:||||: |||||
1206 AAAAAAGAAATCAGGCTCTGAAAGAGAGGCGAGATTCTCTGCCCAAGGAG 1255
359 IleLysTyrlMetLysAsnLeuMetGluAspValCysLysAlaLysGly 374
::: ||||:||||:||||:||||:||||:||||: ||||| |||||
1256 ATCCAGTATCTGAAAGACCTGTAGAAGAGTCCGTAAGGCAAGGGG 1303

```

seq_name: gb_ov:AB013138

seq_documentation_block:

LOCUS AB013138 1461 bp mRNA VRT 02-APR-1999
DEFINITION Gallus gallus mRNA for ATP4, complete cds.

ACCESSION AB013138

VERSION AB013138.1 GI:4586686

KEYWORDS ATP4.

SOURCE Gallus gallus 3.5 day old eye cDNA to mRNA.

ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;

Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE 1 (sites)

Murata,T., Suzuki,H., Ogino,H. and Yasuda,K.

Isolation and analysis of embryonic expression pattern of chicken

ATF4

JOURNAL Unpublished (1998)

REFERENCE 2 (bases 1 to 1461)

AUTHORS Murata,T.

TITLE Direct Submission

JOURNAL Submitted (21-APR-1998) to the DDBJ/EMBL/GenBank databases. Takuya

Murata, Nara Institute of Science and Technology, Graduated

School of Biological Science; 8916-5 Takayama-cho, Ikoma, Nara

630-0101, Japan (E-mail:t-murata@bs.ait-nara.ac.jp,

Tel:++81-743-72-5550, Fax:++81-743-72-5559)

FEATURES

source

1. .1461

/organism="Gallus gallus"

TITLE	Major, T., McFarland, J., Newton, J., Osborne-Lawrence, S., Schageman, J., Schultz, R.A., Stimson, S., Syed, M. and Ward, T.
REFERENCE	HTGS Submission
AUTHORS	2 (bases 1 to 80920)
	Evans, G.A., Bradbury, P., Brignac, S., Bumeister, R., Burbee, D., Davies, G.A., Davis, C.J., Davis, C., English, C., Fondon, T., Franklin, T.L., Garner, H.R., Gordon, M., Gotway, G., Grant, O., Hahner, L., Harris, J., Hinson, S., Megarity, C., Narayanaswamy, U., Newton, J., O'Brien, K., Oliver, T., Patel, P., Probst, S., Rayner, S., Schageman, J., Schilling, P., Schultz, R., Syed, M., Valenzuela, D., Ward, T. and Wilson, R.
TITLE	Direct Submission
JOURNAL	Submitted (24-APR-1997) Genome Science and Technology Center, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA
REFERENCE	3 (bases 1 to 80920)
AUTHORS	Evans, G.A., Athanasios, M., Basit, M., Bradbury, P., Brignac, S., Bumesiter, R., Davis, C., English, C., Franklin, T.L., Garner, H.R., Gee, V., Gordon, M., Gotway, G., Grant, O., Hahner, L., Harris, J., Hinson, S., Narayanaswamy, U., Newton, J., O'Brien, K., Patel, P., Schageman, J., Schilling, P., Schultz, R., Syed, M., Valenzuela, D., Ward, T. and Wilson, R.
TITLE	Direct Submission
JOURNAL	Submitted (17-DEC-1997) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA
REFERENCE	4 (bases 1 to 80920)
AUTHORS	Evans, G.A., Athanasios, M., Aguayo, P., Armstrong, D., Basit, M., Buettner, J., Butler, C., Card, P., Desailboat, F., Dunn, J., English, C., Ethridge, S., Garner, H.R., Gee, V., Gordon, M., Gotway, G., Grant, O., Hahner, L., Joslin, J., Lewis, E., Loo, H., Loo, K.N., Major, T., McFarland, J., Newton, J., Osborne-Lawrence, S., Schageman, J., Schultz, R.A., Stimson, S., Waller, K. and Ward, T.
TITLE	Direct Submission
JOURNAL	Submitted (30-OCT-1998) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA
COMMENT	On Oct 30, 1998 this sequence version replaced gi:2695562. IMPORTANT: This submission contains the entire insert of clone pDJ356d6. pDJ356d6 comes from the RPCI-3 PAC library constructed at the Roswell Park Cancer Institute by the Pieter de Jong group. This clone has been finished according to strict quality criteria and attempts have been made to resolve all base calling problems such as compressions and repetitive elements. The expected Phred/Phrap calculated errors/10kb is 0.49. In addition, attempts have been made to assure 99% of consensus base calls consist of either double-stranded coverage or 2 types of labeling chemistry on one strand.
FEATURES	Location/Qualifiers
source	1. 80920
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
repeat_region	complement(2..356)
repeat_region	/rpt_family="Alu"
repeat_region	complement(437..723)
	/rpt_family="Alu"
	1892..2123
repeat_region	/rpt_family="Alu"
repeat_region	complement(2426..3043)
	/rpt_family="L1"
repeat_region	complement(3085..3379)
	/rpt_family="Alu"
repeat_region	4272..4528
	/rpt_family="Alu"
repeat_region	4543..5068
	/rpt_family="ML2B2"
repeat_region	complement(5137..6056)
	/rpt_family="Alu"
repeat_region	complement(6085..6628)
	/rpt_family="L1"
repeat_region	complement(7044..7319)
	/rpt_family="Alu"

```

/rpt_family="Alu"
complement(50507..50800)
/rpt_family="Alu"
50961..51248
/rpt_family="Alu"
complement(52857..53155)
/rpt_family="Alu"
complement(54353..54620)
/rpt_family="Alu"
complement(57690..57970)
/rpt_family="Alu"
58440..58712
/rpt_family="Alu"
complement(59346..59647)
/rpt_family="Alu"
59759..59911
/rpt_family="MER46"
complement(59761..60003)
/rpt_family="MER46"
62049..62321
/rpt_family="Alu"
63007..63467
/rpt_family="Alu"
63853..64138
/rpt_family="Alu"
68266..68555
/rpt_family="Alu"
68837..69388
/rpt_family="Alu"
complement(69499..69834)
/rpt_family="MER30"
complement(70959..71262)
/rpt_family="Alu"
71879..72057
/rpt_family="MER8"
complement(72001..72057)
/rpt_family="Tigger2"
complement(72691..72979)
/rpt_family="Alu"

alignment_scores:
  Quality: 231.00      Length: 386
  Ratio: 1.111        Gaps: 18
  Percent Similarity: 53.886  Percent Identity: 29.793

alignment_block:
US-08-656-811A-1 x AC002036/rev ..

Align seg 1/1 to reverse of: AC002036 from: 1 to: 80920

19  GluMetProValGlnThrAspGlyGlnPheGlyAspLeuLysSer.. 34
||||| ..:::.....:|||||.....:|||||
79069  GAATGACCTTCCAGGACGAGGATATGTTGGGGACTTGATGTCCTCC 79020
35  .....ThrSerArgHisGlyGlyAspGluSerLeu..... 46
||||| ..:::.....:|||||.....:|||||
79019  CTTGCACGATCGGTTTGGGGCTGAAGAAAGCCTAGCTCTCTAGATA 78970
47  .....:.....:GlnProGlnGlyAlaThrLeuLys 54
||||| ..:::.....:|||||.....:|||||
78969  ACTACCTGGAGGTGCCAAGCAGCTTCAACCTCATGGG..... 78932
55  LeuGluProPheGluGluAspValLeuGlyAla.....GluTpmE 68
||||| ..:::.....:|||||.....:|||||
78931  .....TTCTCAGGACAAGCTTAAGCGGGGCTTCTCCGAATGGCT 78891
68  tGluSerSerAspLeuGly.....SerPheLeuAspAla 80
: ..:::.....:|||||.....:|||||
78890  GCGTGTGGATGGTTAGGCAGTCCCTCCAAATACAGCAAGGAGATGCT 78841
80  euGlyAspAsnHisGluArgLeuHisProPheGluSerAsnLeuGlu 96
||||| ..:::.....:|||||.....:|||||
78840  TCTCCGGACAGATTTGGATGTTGGAGAAATGGATTTGAAGGAGTTCGTC 78791
||||| ..:::.....:|||||.....:|||||

```

```

seq_name: gb_ro:AB012277

seq_documentation_block:
LOCUS      AB012277             880 bp      mRNA                      25-AUG-1999
DEFINITION Mus musculus mRNA for ATF4, partial cds.
ACCESSION  AB012277
VERSION    AB012277.1  GI:2988374
KEYWORDS   ATF4.
SOURCE     Mus musculus
ORGANISM   Mus musculus cDNA to mRNA.
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS    Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE      Nishizawa, M. and Nagata, S.
JOURNAL    Nishizawa, M. and Nagata, S.
MEDLINE    cDNA clones encoding leucine-zipper proteins which interact with
REFERENCE  G-CSF gene promoter element 1-binding protein
AUTHORS    FEBS Lett. 299 (1), 36-38 (1992)
TITLE      92183900
JOURNAL    2 (bases 1 to 880)
MEDLINE    Nishizawa, M. and Nagata, S.
REFERENCE  Direct Submission
AUTHORS    Submitted (17-MAR-1998) to the DDBJ/EMBL/GenBank databases, Mikio
TITLE      Nishizawa, Kansai Medical University, Department of Medical
JOURNAL    Nishizawa, 10-15 Fumizono, Moriguchi, Osaka 570-8506, Japan
MEDLINE    (E-mail:nishizawa@cakii.kmu.ac.jp, Tel:81-6-992-1001(ex.2454),
REFERENCE  Fax:81-6-992-1781)
FEATURES             Location/Qualifiers
     source            1..880
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     CDS               <1..792
                     /codon_start=1
                     /product="ATF4"
                     /protein_id="BAA25314.1"
                     /db_xref="GI:2988375"
                     /translation="KMDLKEFFDALFRMDDLETWPELLTLDDTCDLFAPLVQETIN
                     KEPQVNVIGHLSLVKDVQVPTFLQPPCSPGVLSSTPEHFSLELGSVEVDIS
                     EDGRPDSAAYITLIPPCVKEEDTSDNDSGICMSPESYIGSPHSPTSRAPDNLPL
                     SDRKSPGPRPYDPYDPPKVKTEKLDKLLKKMEQNTAATRYRQKRAEQEAL
                     TGECKELEKNEALKEKADSLAKEIQYLDLITEVRKARKKRPV"
     polyA_signal      849..854
     BASE COUNT       246 a 207 c 218 g 209 t
     ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859

```



```

344 u$anThrLysLeuLysThrLysValAspAspLeuGlnArgGluLeuLysT 361
    ||| :|||:|||||
555 CAATCCGAGCTCAAGGAACGGCGACAGTCCGTCGAGCGCGAGATCCAGT 604
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
361 y$MeLysAsnLeuMetGluAspValCysLysAlaLys 373
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
605 ACGTCAAGGACCTGCATCGAGGTTTACAAGGCCCGG 642

seq_name: gb_htg24:AC011452

seq_documentation_block:
LOCUS      AC011452 147307 bp      DNA               HTG               06-APR-2000
DEFINITION Homo sapiens chromosome 19 clone CTC-326K19, WORKING DRAFT
SEQUENCE, 3 unordered pieces.
ACCESSION  AC011452
VERSION    AC011452.3 GI:7458719
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 147307)
            DOE Joint Genome Institute.
            Sequencing of Human Chromosome 19
            Unpublished
            2 (bases 1 to 147307)
            DOE Joint Genome Institute.
            Direct Submission.
            Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            On Apr 6, 2000 this sequence version replaced gi:6604391.
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----Summary Statistics
            Consensus quality: 145829 bases at least Q40
            Consensus quality: 146455 bases at least Q30
            Consensus quality: 146724 bases at least Q20
            Estimated insert size: 147307; sum-of-contigs estimation
            Estimated insert size: 147530; agarose-fp estimation
            Quality coverage: 6.40x in Q20 bases; agarose-fp estimation
            Quality coverage: 6.41x in Q20 bases; sum-of-contigs estimation
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            1 22842: contig of 22842 bp in length
               gap of unknown length
            * 22843 81048: contig of 58206 bp in length
               gap of unknown length
            * 81049 147307: contig of 66259 bp in length.
               Location/Qualifiers
            1. 147307
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="19"
               /clone="CTC-326K19"
BASE COUNT  36775 a 37696 c 38022 g 34814 t
ORIGIN

alignment_scores:
Quality: 188.50      Length: 146
Ratio: 2.094         Gaps: 2
Percent Similarity: 61.644      Percent Identity: 30.822

```

```
alignment_block:
US-08-656-811A-1 x AC011452  ..
Align seg 1/1 to: AC011452 from: 1 to: 147307

234 ProSerSerProGluThrSerGlnSerSerIlelleGluSerSerProGlu 250
|||||  |||  ::|  |||  ::|  ::|  |||
120204 CCTCCCTCCCTCCCTCCCTCCCTCCCTTTGACCTCCGCCACGCCCTGT 120253
250 uLeuTyrLysValIleSerThrSerSerIleAspAlaSerLysArgPheS 267
|||  |||  ::|  |||  ::|  |||  ::|  |||
120254 CTG.....GATACTCTGGACTGTGCTGG 120276
267 erProTyrSerArgSerSerLysSerLysGlnSerValLysThrSerAsp 283
::|  |||  ::|  ::|  ::|  ::|  ::|  |||
120277 CCATCTACTGCCGCAACGAGCGCGGAGGAGGAAGTGGGGATGCCGCT 120326
284 AlalysAlaProArgLysThrArgThrProAlaGlnProValProGlu.. 299
|||  ::|  |||  ::|  |||  |||  |||  |||
120327 CTGCCCCCGCCACAGCAGCCCTCTCTCTTCCACCTCAACCTTCTCG 120376
300 .....HisValIleMetGluHisLeuAspLysLysAspA 311
|||  ::|  ::|  ::|  ::|  ::|  |||  |||
120377 CTTGCCCTTACCACATCTGTCACACCCGAGGGGACCGCAAGCAAA 120426
311 rGlyLysLeuGlnAsnLysAsnAlaIleArgTyrArgMetLysLys 327
::|  ::|  |||  |||  ::|  |||  ::|  |||  ::|  |||
120427 AGAAGAGAGACAGACAAGTCGGCGGCTCTGAGGTACCGCCAGCGGAAG 120476
328 LysGlyGluAlaGlnGlyIleLysGlyGluGlnGluLeuGluLe 344
::|  ::|  ::|  ::|  ::|  ::|  ::|  |||  |||  |||
120477 CGGCGAGAGGTGAGGCCCTGGAGGGCGAGTCCAGGGGCTGGAGGCACG 120526
344 uAsnThrLysLeuLysThrLysValAspLysLeuGlnArgGluIleLysT 361
|||  ::|  |||  |||  ::|  ::|  ::|  ::|  ::|  |||  |||
120527 GAATCGCGAGCTGAGGAAACCGGAGAGTCCGTGGAGCGGAGATCCAGT 120576
361 yrMetLysAsnLeuMetGluAspValCysLysAlaLys 373
|||  ::|  |||  ::|  |||  ::|  |||  |||  |||
120577 ACGTCAAGGACCTGCTCATCGAGTTTACAGGCCCGG 120614
```

OM of: US-08-656-811A-1 to: N_Geneseq_36.* out_format : pfs

Date: Aug 8, 2000 8:43 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame-pzn.model -DEV=xlp
-Q/cgn2_1/USPTO_spool/US08656811/runat_25072000_142318_17131/app_query.fasta_1.443
-DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -MINLEN=0 -MAXLEN=100000 -USER=US08656811_@Cgn1_1_88
-NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-08-656-811A-1

Query length: 379

Database: N_Geneseq_36.*

Database sequences: 311585

Database length: 125096042

Search time (sec): 54.050000

score_list:

Sequence	Strd Orig	ZScore	EScore Len	Documentation
N_Geneseq_36:V04079	+ 1923.00	2891.57	1.8e-153	1336
N_Geneseq_36:T77784	+ 137.50	199.36	0.0017	669
N_Geneseq_36:V29204	+ 132.00	175.13	0.0373	3190
N_Geneseq_36:T14025	+ 128.50	183.96	0.0120	798
N_Geneseq_36:V69717	+ 124.50	161.41	0.2169	4031
N_Geneseq_36:V69720	+ 124.50	160.93	0.2307	4225
N_Geneseq_36:T17502	+ 121.50	152.97	0.6404	5914
N_Geneseq_36:Q66090	+ 120.50	169.52	0.0767	1005
N_Geneseq_36:Q13674	+ 120.50	156.75	0.3944	3518
N_Geneseq_36:T17471	+ 120.50	151.46	0.7775	5914
N_Geneseq_36:T17506	+ 120.50	151.46	0.7775	5914
N_Geneseq_36:V32452	+ 119.50	145.69	1.63	8982
N_Geneseq_36:T42021	+ 118.50	148.89	1.08	5656
N_Geneseq_36:T70064	+ 118.50	148.89	1.08	5656
N_Geneseq_36:V60560	+ 118.50	148.89	1.08	5656
N_Geneseq_36:T42023	+ 118.50	148.83	1.09	5689
N_Geneseq_36:T70066	+ 118.50	148.83	1.09	5689
N_Geneseq_36:V60562	+ 118.50	148.83	1.09	5689
N_Geneseq_36:T42031	+ 118.50	148.80	1.09	5707
N_Geneseq_36:T70074	+ 118.50	148.80	1.09	5707
N_Geneseq_36:V60570	+ 118.50	148.80	1.09	5707
N_Geneseq_36:T42027	+ 118.50	148.80	1.09	5709
N_Geneseq_36:T42022	+ 118.50	148.80	1.09	5709
N_Geneseq_36:T70065	+ 118.50	148.80	1.09	5709
N_Geneseq_36:V60569	+ 118.50	148.80	1.09	5709
N_Geneseq_36:T70070	+ 118.50	148.80	1.09	5709
N_Geneseq_36:V171738	+ 118.50	148.80	1.09	5709
N_Geneseq_36:V60561	+ 118.50	148.80	1.09	5709
N_Geneseq_36:V60566	+ 118.50	148.80	1.09	5709
N_Geneseq_36:T42026	+ 118.50	148.79	1.09	5710
N_Geneseq_36:T70069	+ 118.50	148.79	1.09	5710
N_Geneseq_36:V60565	+ 118.50	148.79	1.09	5710
N_Geneseq_36:T42024	+ 118.50	148.79	1.09	5711
N_Geneseq_36:T70067	+ 118.50	148.79	1.09	5711
N_Geneseq_36:V171737	+ 118.50	148.79	1.09	5711
N_Geneseq_36:V46463	+ 118.50	148.79	1.09	5711
N_Geneseq_36:V46464	+ 118.50	148.79	1.09	5711
N_Geneseq_36:V46449	+ 118.50	148.79	1.09	5711
N_Geneseq_36:V46458	+ 118.50	148.79	1.09	5711
N_Geneseq_36:V46459	+ 118.50	148.79	1.09	5711
N_Geneseq_36:V46460	+ 118.50	148.79	1.09	5711
N_Geneseq_36:V60563	+ 118.50	148.79	1.09	5711

N_Geneseq_36:Q93027 + 118.50 148.79 1.09 5712 ! Gene encoding BRCA1 protein
N_Geneseq_36:T42032 + 118.50 148.79 1.09 5712 ! BRCA1 allele #77. New mutant
N_Geneseq_36:T70075 + 118.50 148.79 1.09 5712 ! Mutant BRCA1 allele #77. BR
N_Geneseq_36:T84840 + 118.50 148.79 1.09 5712 ! Human breast and ovarian ca

seq_name: N_Geneseq_36:V04079

seq_documentation_block:

ID V04079 standard; DNA; 1336 BP.

AC V04079;

DT 22-JUN-1998 (first entry)

DE Aplysia cAMP-response element binding protein 2 DNA.

KW ApCREB-2; cAMP-response element binding protein-2; snail;

KW transcription factor; memory loss; Alzheimer's disease; amnesia;

KW ischaemia; head trauma; neuronal injury; Parkinson's disease;

senility; therapy; ss.

OS Aplysia californica.

Key Location/Qualifiers

FT CDS 190..1326

FT WO9746257-A1.

PD 11-DEC-1997.

PF 03-JUN-1997; U09438

PR 03-JUN-1996; US-658811.

PA (YOCO) UNIV COLUMBIA NEW YORK.

PI Bartsch D, Ghirardi M, Kandel ER;

DR WPI; 98-051903/05.

DR P-PSDB; W41508.

PT Enhancing long-term memory in subjects whose cAMP-responsive gene is

PT repressed - used to treat long-term memory defects, e.g. age-related

PT memory loss, Alzheimer's disease

PS Example 2; Page 73-74; 100pp; English.

CC This polynucleotide includes a coding region for the cAMP-response

CC element binding protein-2 (see W41508), i.e. ApCREB-2, of the

CC marine snail Aplysia. ApCREB-2 is a transcription factor and

CC repressor of long-term facilitation in Aplysia neurons. It is a

CC homologue of human CREB-2 and mouse Atf-4. The polynucleotide

CC sequence was deduced from 2 independent clones isolated by a yeast

CC two-hybrid screen of an Aplysia central nervous system cDNA library.

CC The invention provides a method of enhancing long-term memory in a

CC subject whose cAMP-responsive gene expression is repressed due to

CC binding of CREB-2 to a protein or DNA associated with cAMP-responsive

CC gene expression. The method involves administering to the subject a

CC compound capable of interfering with such binding so as to derepress

CC cAMP-responsive gene expression and thereby enhance long-term memory.

CC Such compounds include anti-CREB-2 antibodies or a compound capable of

CC altering phosphorylation of CREB-2. The method is used to treat e.g.

CC age-related memory loss, Alzheimer's disease, amnesia, ischaemia,

CC shock, head trauma, neuronal injury, toxicity or degradation,

CC Parkinson's disease or senility (claimed).

CC Sequence 1336 BP; 375 A; 307 C; 328 G; 326 T;

alignment_scores:

Quality: 1923.00 Length: 378

Ratio: 5.087 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-08-656-811A-1 x V04079 ..

Align seg 1/1 to: V04079 from: 1 to: 1336

1 MetGluLeuAspLeuTrpSerGluAspPheGlnLeuAlaArgGluTrpGcl 17

|||||

190 ATGGAGCTGGACCTTTGGAGCGAAGATTTTCACTGGCCAGGAATGGG 239

|||||

17 yLeuGluMetProValValGlnThrAspGlyGlnPheGlyAspLeuLys 34

|||||

240 GCTGGAATGCCAGTCGTCAGACCGATGCCAGTTCGGTGACCTCAAT 289

|||||

34 erThrSerArgHisGlyGlyAspGluSerLeuSerLeuGlnProGlnGly 50

|||||

```
290 CAACACAGTCGTCATGGTGGCGACGAATCTCTAAGTTTCGACGCCCGCAGGC 339
51 AlaThrLeuLysLeuGluProPheGluGluAspValLeuGlyAlaGluTr 67
|||||
340 GCTACACTGAAGTTGGAACCTTTGAGGAGAGATGCTCTTGGTCGAGAGTG 389
67 pMetGluSerSerAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnH 84
|||||
390 GATGAGTCGTCGCGATCTCGGCTCTTTCTGGATGCTTTGGTGGACAAAC 439
84 IsGluArgLeuHisProPheGluSerAsnLeuLeuGluPheThrSerLeu 100
|||||
440 ATGAGCGGCTCATCCGCTCGAGTCAAACTTTCGAGTTCACCTCTCTG 489
101 IleThrProAspAspSerThrValSerLysAspIleLeuSerSerThrLe 117
|||||
490 ATCACTCTGATGATTCGACCGGTGTCAAAGGACATCTCTCACTCAACTCT 539
117 uGlnPheProThrGlnProValAsnIleProLeuTyrAlaSerHisGlyA 134
|||||
540 TCAGTTTCCAACTCAACCACTGAACATCCCTTTATATGCAAGTCATGGGG 589
134 laGluAspPheSerAlaGluThrGluPheGluAsnHisLeuSerProPro 150
|||||
590 CCGAAGATTTCTCTGCGAGAGACTGAGTTTGAGAACCACTCTCGCCTCCA 639
151 AspSerProGluGlnValAlaProValIleAsnLeuGluProValGluLe 167
|||||
640 GATTCCTCGGAGCAGTAGCCCTGTCTAATACTAGAACCACTGTAACCT 689
167 uThrAlaSerHisMetThrValIleSerProAspGlyLeuLeuGlyGlyM 184
|||||
690 CACTGCGAGCCATATGACGGTGATCTCACTACCTGATGGCTTGTGGTGCA 739
184 etGluLeuAlaSerGluSerLeuThrPheThrGluLeuAspPheValAsn 200
740 TCGAACTGGCTTTCAGAAAGCTTAACATTTACCGAACTAGACTTTGTGAAC 789
201 PheAsnAspSerAlaValGlySerIleGlyAlaGluGluLeuGlu 217
|||||
790 TTCATGNCAGTGTGTGGTTCAATTTGGCGGTGCTGAAGAACTTCTTGG 839
217 ySerProLeuSerValAspValGluSerThrIleSerPheSerGlyP 234
|||||
840 CTCGCCACTGTGCTGATGATGTGGAAGTACAATATCATATTTTCAGGTC 889
234 roSerSerProGluThrSerGlnSerSerIleLeuSerSerProGlu 250
890 CATCGTCCGCAACCAACCGCAGACGACATCATTTGAATCAAGTCTTGAA 939
251 LeuTyrLysValIleSerThrSerSerIleAspAlaSerLysArgPheSe 267
940 TTGTACAAAGTTATCTTACTCTGCTCCATTCATTCATCTAAGCGTTTCTC 989
267 rProTyrSerArgSerSerLysSerLysGlnSerValLysThrSerAspA 284
990 TCCATACTCTGTTCTCCCAAGTCCAAGCAATCTGTCAAGACTTCAGAGC 1039
284 laLysAlaProArgLysThrArgThrProAlaGlnProValProGluHis 300
1040 CTAAGGCACCTCGTAAACAGGAGACACCGCGCGACCTGTGCCAGAACAT 1089
301 ValIleMetGluHisLeuAspLysLysAspArgLysLysLeuGlnAsnLy 317
1090 GTCATCATGGAAACATTGGACAAAAGGACAGAAAGAGCTTCAGAACAA 1139
317 sAsnAlaIleArgTyrArgMetLysLysLysGlyGluAlaGlnGlyI 334
1140 GAATGCTGCCATTAGGTATAGGTATAGTAAGAAGAGGGAGGCTCAGGSCA 1189
334 leLysGlyGluGlnGluLeuGluLeuAsnThrLysLeuLysThr 350
1190 TCAAGGGGAGGAAACAGAAATTAGAAGAACTCAACACAAAGCTTTAAGACT 1239
```

```
351 LysValAspAspLeuGlnArgGluIleLysTyrMetLysAsnLeuMetG1 367
|||||
1240 AAGTCGATGACTTGCAAGAGAAATCAAGTACATGAAAAATTTAATGGA 1289
367 uAspValCysLysAlaLysGlyIleGlnLeuLys 378
|||||
1290 AGATGTTTGCAGGCGGAAAGGTATTTCAGCTTAAA 1323
```

seq_name: N_Geneseq_36:T77784

seq_documentation_block:

ID T77784 standard; cDNA; 669 BP.

AC T77784;

DT 01-OCT-1997 (first entry)

DE cDNA clone 59 encoding NIP-1.

NP-1; NIP-2; NuMA; nuclear mitotic apparatus; NIMA interacting protein;

KW cell division; proliferation; antibody; Ab; detection;

KW malignant cell growth; ss.

OS Homo sapiens.

PN WO9640917-A1

PN 19-DEC-1996.

PF 07-JUN-1996; U09504.

PR 07-JUN-1995; US-478408.

PA (UYA) UNIV YALE.

PI McPherson SMG, Snyder MP;

DR WPI: 97-077270/07.

DR P-PSDB; W21733.

PT New nucleic acid encoding nuclear mitotic appts. interacting

PT proteins - useful for modulating cell division and proliferation and

PT in diagnosis

PS Claim 28; Page 57-58; 78pp; English.

CC The sequences given in T77784-86 encode NuMA binding proteins (NuMA -
CC nuclear mitotic apparatus). These protein sequences were identified
CC using the fusion proteins given in W21731-32. Compounds which interfere
CC with the interaction of NuMA with a known NIP (NuMA interacting protein)
CC are used to modulate cell division and/or proliferation. Ab, raised
CC conventionally using NIP-1 or -2 as immunogen, are used to detect NIP
CC (or their complexes) and to block their activity for diagnostic or
CC therapeutic use, e.g. to detect defective NuMA or NIP which may be
CC markers for aberrant (including malignant) cell growth (which can also
CC be detected by nucleic acid sequencing). Also where malignancy is
CC related to defects in NuMA or NIP, it can be treated by administration
CC of the appropriate functional protein. The protein encoded by this
CC sequence is rich in proline residues.
SQ Sequence 669 BP; 131 A; 232 C; 191 G; 115 T;

alignment_scores:

Quality: 137.50 Length: 148

Ratio: 1.599

Percent Similarity: 58.108 Percent Identity: 29.730

alignment_block:

US-08-656-811A-1 x T77784 ..

Align seg 1/1 to: T77784 from: 1 to: 669

234 ProSerSerProGluThrSerGlnSerSerIleLeuGluSerSerProG1 250

28 CCCCCTCCCTCCCTGCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTG 77

250 uLeuTyrLysValIleSerThrSerSerIleAspAlaSerLysArgPheS 267

78 CTGG.....GATCTCTGGACTGCTGG 100

267 erProTyrSerArgSerSerLysSerLysGlnSerValLysThrSerAsp 283

101 CCATCTACTGCGCAACGAGCGCGGAGGAGGAGTGGGATGCCGCT 150

284 AlalysAlaProArgLysThrArgThrProAlaGlnProValProGlu.. 299

151 CTGCCCCCGGCACGACGAGCCCCCTCCCTCTTCACCTCAACCTTCTCG 200


```

300 .....HisValIleMetGluHisLeuAspLysLysAspA 311
      |||      : : : |||:|||||:
201 CTGCCCCCTACCCACATCTCCACCACCGAGGGGACCGCAAGCAA 250
      |||:|||||:
311 tGLYSLYSLeuGlnAsnLysAsnAlaIleArgTyrArgMetLysLys 327
      :|||:|||||:|||||:|||||:|||||:|||||:
251 AGAAGAGACAGAACAAAGTCGGCGGCTCTGAGGTACCGCCAGCGAAGC 300
      :|||:|||||:|||||:|||||:|||||:|||||:
328 tLysGLYGLuAlaGlnGlyIleLysGLYGLuGlnGluGluL 344
      :|||:|||||:|||||:|||||:|||||:|||||:
301 GAGAGGGTGAGCCCTGGA.....GCGAGGTGCCAGGGGTGGAGGCAC 344
      :|||:|||||:
344 euAsnThrLysLeuLysThrLysVal.AsaspLeuGlnArgGluIleLys 360
      ||| :|||:|||||:|||||:|||||:|||||:|||||:
345 GGAATCGCGAGCTGAAGAACGGGCAAGAGTCCGTGGAGCGCGAGATCCA 394
      :|||:|||||:
360 stYrMetLysAsnLeuMetGluAspValCysLysAlaLys 373
      :|||:|||||:|||||:|||||:|||||:
395 GTACGTCAAGGACCTGCTCATCGAGGTTTACAAAGCCCGG 434

```

seq_name: N_Geneseq_36:V29204

seq_documentation_block:

ID V29204 standard; cDNA; 3190 BP.

AC V29204;

DT 10-AUG-1998 (first entry)

DE Gene encoding the mouse cAMP regulatory element binding protein.

KW Mouse cAMP regulatory element binding; mCREBA; antibody; inhibition;

KW mCREBA modulator; ss.

OS Mus sp.

FH Key Location/Qualifiers

FT CDS 304..1869

FT /*tag= a

FT /product= "mCREBA"

FN W09812910-A2.

PD 02-APR-1998.

PF 26-SEP-1997; U17288.

PR 27-SEP-1996; US721684.

PA (ICOS-) ICOS CORP.

PI Keegan KS;

DR WPI; 98-230326/20.

DR P-PSDB; W37934.

PT Murine cAMP regulatory element binding protein A, mCREBA - useful

PT for systematic analysis of structure and function of mCREBA and

PT identification of those molecules with which it will react

PS Claim 3; pages 26-29; 41pp; English.

CC This is the nucleotide sequence of the gene encoding the mouse cAMP

CC regulatory element binding (mCREBA) protein. The products of the

CC mCREBA such as antibodies can be used for the recombinant production

CC of the protein, to identify novel genes encoding binding partner

CC polypeptides for mCREBA, for the systematic analysis of the structure

CC and function of mCREBA, and identification of those molecules with

CC which it will react, to identify inhibitors of mCREBA binding to

CC other natural binding partners and to generate rodents that fail to

CC express a functional mCREBA or express a variant mCREBA, useful as

CC models for studying the activities of mCREBA and mCREBA modulators in

CC vivo.

SQ Sequence 3190 BP; 729 A; 939 C; 778 G; 744 T;

alignment_scores:

Quality: 132.00 Length: 423

Ratio: 0.635 Gaps: 22

Percent similarity: 49.173 Percent identity: 21.513

alignment_block:

US-08-656-811A-1 x V29204

Align seg 1/1 to: V29204 from: 1 to: 3190

19 GluMetProValValGlnThrAspGlyGlnPheGlyAspLeuLysSerTh 35

|||||: |||:||||| ||| :|||:|||||:|||||:|||||:|||||:

```

325 GACGACGAGCGTCTCTGCAGTGGGACCCCAAGCTGTAGCGAGCTG..... 366
35 rSerArgHisGlyGlyAspGluSerLeuGlnProGlnGlyValat 52
   |||:|||||:
367 .TCAGAGCCCGGAGAGACTGAGCCCTCATGTACACACGCACTTCTCGG 415
52 hrLeuLysLeuGluProPheGluGluAspValLeuGlyAlaGluTyrMet 68
   ||| |||:|||||:|||||:|||||:
416 AGCTC...CTAGACGAGTTCCTCCAGAGCTCTCTGGGT..... 450
69 GluSerSerAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisG1 85
   :|||:|||||:
451 .....CAGCTCTCGAGTGAC..... 465
85 uArgLeuHisProPheGluSerAsnLeuGluPheThrSerLeuIlet 102
   |||:|||||:
466 .....CCTTCTCTCTCA..... 477
102 hrProAspAspSerThrValSerLysAspIleLeuSerSerThrLeuGln 118
   :|||:|||||:
478 .....GAGAAGAGCGAGTCAATGGAGGTGGAGCCATCT..... 510
119 PheProThrGlnProValAsnIleProLeuTyrAlaSerHisGlyAlaG1 135
   |||:|||||:|||||:
511 ...CCAACATCACCGCG.....CCTCTCATCCAGGCTGAACACAGCTA 551
135 uAspPheSerAlaGluThrGluPheGluAsn.....HisLeuSerP 149
   :|||:|||||:
552 CTCTCTGAGCGAGGAGCCCGGACTCAGTCACCATTTACCCATCGCGCTA 601
149 roProAspSer.....ProGluGlnVal..... 156
   |||:|||||:
602 CCAGCGACAGCTTCAATGACGAGGAGGTGGAGAGTGAAAAATGGTACCTG 651
157 .....AlaProValIleAsnLeuGluProValGluLeuTh 168
652 TCTACAGAGTTTCTCTTACCTACCTACCATCAAGAAAGACCAATC..... 693
168 rAlaSerHisMetThrValIleSerProAspGlyLeuLeuGlyGlyMetG 185
   ||| :|||:|||||:
694 .....ACAGAGGAGCAGCCCGCGGACTT..... 717
185 luLeuAlaSerGluSerLeuThrPheThrGluLeuAspPheValAsnPhe 201
   :||| :|||:|||||:|||||:
718 ..GTCCCTTCTGTCTGCTGCTGACCATCACAGCCATTTCC...ACTCCTTT 762
202 Asn.....AspSerAlaValGlySerIleGlyAl 212
   :|||:|||||:
763 GAAAAGAAAGAGTCCCTCTGGATATGAATGTGGGGGGGACTCCTCATG 812
212 aGluGluLeuLeuGlySer...ProLeuSerValAspAspValGluSert 228
   :|||:|||||:
813 CCAGAGCGCTTATTCCTAAGATTAAAGCTGGAGCCCGCCACGAAGTGGATCAGT 862
228 hrIleSerPheSerGlyProSerSerProGluThrSerGlnSerSerIle 244
   :|||:|||||:|||||:
863 TCTTAAACTTCTCC...CGAAAGAGCCTCGGTGGATCAACTGCACCTTA 909
245 IleGluSerSerProGluLeuTyrLysValIleSerThrSerSerIleAs 261
   :|||:|||||:
910 CCACCAACACACCCACAGTAGTCACAGCAGTGTCTGTAGGGGGAGCTTGAG 959
261 pAlaSerLysArgPheSerProTyrSerArgSerSer..... 273
960 CCCCACACCCAGCGCTGCTATCCCTTCAGCCTGTCTCAGGCCCCACAGCCCTG 1009
274 .....LysSerLysGlnSerValLysThrSerAsp...Ala 284
1010 TCAGAGCCCATGCCCGGGGCCCTCTGCTGTGTCCACATCTCTCTCTCTC 1059
285 LysAlaProArgLysThrArg..... 291
1060 ACAGTCCACATAAGCTGCAGGGATCGGCCCCCTCTGTCTCTGACAGAAGA 1109

```

```

292 .....ThrpAlaGlnProValProGluHisValI 302
1110 GGAGAAGAGGAGCCCTGCTGGCGAGGCTATCCCATCCCAAGAGCTGC 1159
302 leMet.....GluHisLeuAspLysLysAspArgLysLysLeu 314
1160 CTCTGACAAAATCTGAGGAGAGGCCCTGAAGAAAATCCGGAGAAAGATC 1209
315 GlnAsnLysAsnAlaAlaIleArgTyrArgMetLysLysLysGlyGluAl 331
1210 AAGAAATAGATTCTGCCCAAGAAGCAGGAGAAAGAAAGAAAGAAATACAT 1259
331 aglnGlyLysLysGlyGluGluGlnGluLeuGluGluLeuAsnThrLysL 348
1260 GGACAGCCTGGAGAAAAGTGGAGTCTGTTCACACTCAGAACTTGGAGC 1309
348 euLysThrLysValAspLysLeu.....MetGluAspValCy 370
1310 TTCGGAAGAGGTGGAGTGTCTGGAGAACACCAATAGGACTCTCTCTTCA 1359
358 GluIleLysTyrMetLysAsnLeu.....MetGluAspValCy 370
1360 CAACCTTCAGAGCTTCAGACTTTGGTGATGGGGAAGGTCTCTCGAAGCTG 1409
370 sLysAlaLysGlyIleGln 376
1410 CAAGTTAGCTGGCACACAG 1428
seq_name: N_Geneseq_36:T14025
seq_documentation_block:
ID T14025 standard; cDNA; 798 BP.
AC T14025;
DT 09-JUL-1996 (first entry)
DE Drosophila dCREB1 cDNA.
KW Enhancer-specific activator; dCREB1; long-term memory;
KW transgenic animal; insect; ds.
OS Drosophila sp.
PN W09611270-Al.
PD 18-APR-1996.
PF 06-OCT-1995; U13198.
PR 07-OCT-1994; US-319866.
PR 21-DEC-1994; US-361063.
PA (COLD-) COLD SPRING HARBOR LAB.
PI Regulski M, Tully TP, Yin JC;
DR WPI; 96-209851/21.
DR P-PSDB; R91295.
PT Novel cyclic 3',5'-adenosine monophosphate responsive
PT transcriptional activator gene - used in stimulation and enhancement
PT of longterm memory
PS Claim 31: Page 100-101; 160pp; English.
CC The Drosophila CREB1 gene (T14025) codes for an enhancer-specific
CC activator (R91295) that appears to be a member of the CREB/ATF
CC family and that is associated with long-term memory. The gene was
CC isolated by a DNA-binding expression screen of a Drosophila head cDNA
CC library using a probe contg. 3 CAMP-responsive element sites. dCREB1
CC and dCREB2 (see also T14024) genes can be used to manipulate memory,
CC and transgenic insects carrying the genes can be used to assess the
CC effects of drugs on long-term memory formation.
SQ Sequence 798 BP; 245 A; 173 C; 175 G; 205 T;

```

alignment_scores:

Quality: 128.50 Length: 310
Ratio: 0.824 Gaps: 18
Percent Similarity: 50.323 Percent Identity: 24.516

alignment_block:

US-08-656-811a-1 x T14025

Align seg 1/1 to: T14025 from: 1 to: 798

```

91 GluSerAsnLeuLeuGluPheThrSerLeuIleThrProAspAspSerTh 107
13 GAAAATATGTTTCTACTTTTCACATCGTTA.....GATGCTGCTAC 53
107 rValSerLys..... 110
54 CGGTACCAACCAACACCGGTGAATTTCTTAATGAATGAATCTCCAAGGCAAG 103
111 .....AspIleLeuSerSerThrLeuGlnPheProThrGlnProVal 124
104 AAGCCGGTGACTTAATGTTGAGTACTCTGGATTTCAACATTTATGCGGCAA 153
125 AsnIleProLeuTyrAlaSerHisGlyAlaGluAspPheSerAlaGluTh 141
154 AACCTG.....GCAGATGATTTCAGACCTCG.. 180
141 rGluPheGluAsnHisLeuSerProProAspSerProGluGlnValAlaP 158
181 .....GCTTCACCAGCTTCGGAGGACAAAGATGACTC 211
158 ro...ValIleAsnLeuGluProValGluLeuThrAlaSerHisMetThr 173
212 CTTTCGTTGTGATACCAATGTTTTTGAATCCGCTTCAAGAAC..... 255
174 ValIleSerProAspGlyLeuLeuGlyMetGluLeuAlaSerGluSe 190
256 .....ACCGAAGATACCTTCTAGGA.....GATATCGACAATGTGG 293
190 rLeuThrPheThrGluLeu...AspPheValAsnPheAsnAspSerAlav 206
294 TATTGTTGACACGGAGTGAAGGAGATGTTGCGATTGTTGACTCGGAAA 343
206 alGlySerIleGlyGlyAlaGluLeuLeuGlySerProLeu..... 220
344 TCAATAAC.....GGCACTCCCTATCAAGCAG 369
221 ...SerValAspValGluSerThrIleSerPheSerGlyProSerSe 236
370 GAACAAAAGGATGATTGGAATTACT.....TC 398
236 rProGluThrSerGlnSerSerIleIleGluSerSerProGluLeuTyrL 253
399 AAGATCCCAAGTCCACCTCAGCTCTCTTGTGCTCG..A 433
253 ysValIleSerThrSerSerIleAspAlaSerLysArg...PheSerPro 268
434 AATCGACTTCTGCTTCTCCAGCTGATGCTGCCGCTGCATGTGCAAGTCT 483
269 TyrSerArgSerSerLys.....SerLysGlnSerValLysThrSe 282
484 TCGTCATCGTCTTGTAAAGATCCTATTCTTCTGCTCAGCTAGAACTAC 533
282 rASpAlaLysAlaProArgLysThrArg.....ThrPro..... 293
534 GGGTTCGATGCTCCAAAGAAAGATAAGCTGGGCTGCACCCCTTACACTA 583
294 .....AlaGlnProValProGluHisValIleMetGluHisLeu 306
584 GAAACACAGAGAAACATCCATTACTCGG...GTCATTCCAAAGGGTCAG 630
307 AspLysLysAspArgLysLysLeuGlnAsnLysAsnAlaIleArgTy 323
631 GATGTTGCTTCTATGAAAGGGCAAGAACACTCAGGCCGCAAGATC 680
323 rArgMetLysLysGlyGluAlaGlnGlyIleLysGlyGluGlnG 340
681 AAGAGCCAGAAAATGGAAGAATG.....TCCC 709
340 luLeuGluGluLeuAsnThrLysLysValAspLysLeuGln 356
710 AACTTGAGAAAGTGTCAAACCTTGTGAAGGAAACGACGACTTGAAA 759
357 ArgGluIleLysTyrMetLysAsnLeuMet 366

```



```

2730 AGTCCTTGATAGAGCGCCCTGTTCACTTATACACTGGATGAAAG 2779
306 LeuAspLysLysAspArgLysLysLeuGlnAsnLysAsnAlaAlaIleAr 322
2780 GTGACGAGTGTGGCGGTTCTTCTC.....CTCAA 2811
322 gTyrArgMetLysLys 327
2812 ATATCAAGTGAAGCAG 2827

```

seq_name: N_Geneseq_36.V69720

seq_documentation_block:

```

ID V69720 standard; cDNA; 4225 BP.
AC V69720;
DT 01-MAR-1999 (first entry)
DE Tumour rejection antigen precursor MAGE-C1 cDNA.
KW MAGE-C1; human; tumour rejection antigen precursor; TRAP;
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 257..3685
FT /*tag= a
FT WO9849184-A1.
PD 05-NOV-1998.
PF 24-APR-1998.
PR 25-APR-1997; US-845528.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Boon-Faleur T, De Smet C, Lucas S;
DR WPI; 99-024041/02.
DR P-PSDB; W81546.
PT Tumour rejection antigen precursors - used for determining presence
PT of cytolytic T cells specific for complexes of a human leukocyte
PT antigen
PS Claim 1; Page 48-50; 84pp; English.
CC This nucleotide sequence encodes novel human tumour rejection
CC antigen precursor (TRAP) MAGE-C1 (see W81546). MAGE-C1 is a novel
CC member of the MAGE family that may be recognised by cytotoxic T
CC cells, leading to lysis of the tumour cells which express it.
CC MAGE-C1 and MAGE-C2 (see W81547) are expressed in a variety of
CC tumours and in normal testis cells, but not by other normal cells.
CC The MAGE-C1 cDNA was isolated from a melanoma LB373-MEL cDNA by
CC PCR amplification (see V69732-33). It shows homology to MAGE-A1
CC cDNA (see V69719). The MAGE-C1 gene was localised to Xq26-q27.
CC MAGE-C1 and MAGE-C2 cDNAs (see V69726) are claimed, as are:
CC expression vectors; transformed or transfected cell lines (e.g. COS
CC and CHO); an isolated TRAP encoded by the cDNAs; a kit useful in a
CC PCR based assay; a method for determining expression of a MAGE-C1
CC gene using the kit; a polypeptide comprising a number of tumour
CC rejection antigens derived from MAGE-C1 or MAGE-C2; and a polypeptide
CC comprising at least one tumour rejection antigen derived from
CC MAGE-C1 or MAGE-C2 and at least one other tumour rejection antigen.
CC MAGE-C1 and MAGE-C2 can be used in a method for determining the
CC presence of cytolytic T cells specific for complexes of a human
CC leukocyte antigen (HLA).
SQ Sequence 4225 BP; 871 A; 1198 C; 923 G; 1233 T;

```

alignment_scores:

```

Quality: 124.50 Length: 372
Ratio: 0.688 Gaps: 17
Percent Similarity: 48.656 Percent Identity: 22.043

```

alignment_block:

```

US-08-656-811A-1 x V69720 ..
Align seg 1/1 to: V69720 from: 1 to: 4225
5 LeuTrpSerGluAspPheGlnLeuAlaArgGluTrpGlyLeuGluMetPr 21
|||||.....
2102 CTTGAGGGGAGGAGTTCAGTCTTCT.....CTCCAGAGCCC 2139

```

```

21 oValValGlnThrAspGlyGlnPheGlyAspLeuLysSerThrSerArgH 38
|||||.....TCCTCCACTCCATCCAGTCTTCCCC 2180
2140 TGTGAGCATCTGCTCC.....TCCTCCACTCCATCCAGTCTTCCCC 2180
38 isGlyGlyAspGlySerLeuSerLeuGlnProGlnGlyAlaThrLeuLys 54
|||||.....TCCTCCACTCCATCCAGTCTTCCCC 2180
2181 AGAGTTTCCCTGAGAGTTCCTCAGAGTCTCTGAGGGGCTGTC..... 2224
55 LeuGluProPheGluGluAspValLeuGlyAlaGluTrpMetGluSerSe 71
|||||.....TCCTCCACTCCATCCAGTCTTCCCC 2180
2225 CAGTCTCTCTCCATGATCTCCTCAGAGCCCTCCTGAGGGGATGCACTCCA 2274
71 rasPLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeuH 88
|||||.....TCCTCCACTCCATCCAGTCTTCCCC 2180
2275 ATCTCTCTCTCCAGAGTCTCTGAGAGTCTCTGAGGGGAGGATTCCTGT 2324
88 isPropGluSerAsnLeuGluPheThrSerLeuLeuThrProasp 104
|||||.....TCCTCCACTCCATCCAGTCTTCCCC 2180
2325 CTCCTCTCCAA.....ATTCTCTCAG 2344
105 AspSerThrValSerLysAspIleLeuSerSerThrLeuGlnPhePro.. 120
|||||.....TCCTCCACTCCATCCAGTCTTCCCC 2180
2345 AGTCTCTCTGAGGGAGAGGACTCCCTGTCTCT...CTCCATTTCTCTCA 2391
121 .....ThrGlnProValAsnIleProLeuT 129
|||||.....TCCTCCACTCCATCCAGTCTTCCCC 2180
2392 GAGTCTCTCTGAGTGGGAGGACTCCCTCTCTCTCTCCACTTCTCTCAGT 2441
129 yrAlaSerHisGlyAlaGluasphe..... 137
|||||.....TCCTCCACTCCATCCAGTCTTCCCC 2180
2442 TTCTCTCTCTCAGGGG...GAGGACTTCCAGTCTCTCTCCAGAGTCTGTG 2488
138 .....SerAlaGluThrGluPhe.....G1 144
|||||.....TCCTCCACTCCATCCAGTCTTCCCC 2180
2489 AGTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2538
144 uAsnHisLeuSerProProAspSerProGluGlnValAlaProValIleA 161
|||||.....TCCTCCACTCCATCCAGTCTTCCCC 2180
2539 GAGTCTCTCAGAGTCTCTCTGAGGGGCTGCTCAG...TCTCTCTCTCT 2581
161 snLeuGluProValGlu.....Leu 167
|||||.....TCCTCCACTCCATCCAGTCTTCCCC 2180
2582 ..CAGAGAGCTCTCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2629
168 ThrAlaSerHisMetThrValIleSerProAspGlyLeuLeuGlyGlyMe 184
|||||.....TCCTCCACTCCATCCAGTCTTCCCC 2180
2630 CAAGTTCCTCAGAGTCTCTCTGAGGGGCTCTCTCTCTCTCTCTCTCT 2665
184 tGluLeuAlaSerGluSerLeuThrPheThrGluLeuAspPheValAsn 201
|||||.....TCCTCCACTCCATCCAGTCTTCCCC 2180
2665 ..... 2665
201 heAsnAspSerAlaValGlySerIleGlyAlaGluGluLeuLeuGly 217
|||||.....TCCTCCACTCCATCCAGTCTTCCCC 2180
2666 .....GGGCTCTGCCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2689
218 SerProLeuSer.....ValAspAspValGluSerThrIleSerPheSe 232
|||||.....TCCTCCACTCCATCCAGTCTTCCCC 2180
2690 AGTCTCTGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2739
232 rGlyPro.....SerSerProGluThrSerGlnSerSerIleIleGlu 247
|||||.....TCCTCCACTCCATCCAGTCTTCCCC 2180
2740 TTCCTCTGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2789
247 erSerProGluLeuTyrLysValIleSerThrSerSerIleAspAlaSer 263
|||||.....TCCTCCACTCCATCCAGTCTTCCCC 2180
2790 GTTCCCTGAGAGTCTCTCTCCAGAGTCTCTGTGATCTCTCTCTCTCTCT 2839
264 LysArgPheSerProTyrSerArgSerSer..... 273
|||||.....TCCTCCACTCCATCCAGTCTTCCCC 2180
2840 ACTTCATTCAGCCCATTCAGTTCAGAGAGTCCAGAGCCAGCCAGTAGATA 2889
274 ....LysSerLysGlnSerValLysThrSerAspAlaLysAlaProArgL 289

```

```

2890 TACAAGTTCCTCAGACACCTTCAGAGTGATTCCTTGACGACGCG 2939
      :::::::::::::::::::: |||||::: :::
289 yThrArgThrProAlaGlnProValGluHisValIleMetGluHis 305
      ::::: ::::::::::::::: ::::::::::: |||
2940 AGTCCTTGATAGAGCGAGCCCTGTTGCATTATACACATGGATGAAGA 2989
      ::::::::::::::: |||
306 LeuAspLysLysAspArgLysLysLeuGlnAsnLysAsnAlaAlaIle 322
      ::::::::::::::: |||
2990 GTGAGCGAGTGGCGGCTTCTCTC.....CTCAA 3021
      :::::
322 gTyrArgMetLysLys 327
      ::::::::::::::: |||
3022 ATATCAAGTGAAGCAG 3037
      ::::::::::::::: |||

seq_name: N_Geneseq_36:T17502

seq_documentation_block:
ID T17502 standard; cDNA; 5914 BP.
AC T17502:
DT 02-OCT-1996 (first entry)
DE Mutated BRCA1 coding sequence from PM20.
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 120..5711
FT CDS /*tag= a
FT product= BRCA1 mutant
FT mutation 3238
FT /*tag= b
FT /*note= "G to A mutation"

WO9605306-A2.
PD 22-FEB-1996.
PF 11-AUG-1995; U10202.
PR 12-AUG-1994; US-289221.
PR 02-SEP-1994; US-300266.
PR 16-SEP-1994; US-308104.
PR 29-NOV-1994; US-348824.
PR 24-MAR-1995; US-409305.
PR 07-JUN-1995; US-483553.
PR 07-JUN-1995; US-480784.
PA (MYRI-) MYRIAD GENETICS INC.
PA (CANC-) CANCER INST.
PA (RECH-) CENT RECH DU CHUL.
PI Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;
DR WPI: 96-139702/14.
DR P-PSDB; R81540.
DT New nucleic acid and polypeptide for mutant or polymorphic BRCA1
PT gene - for diagnosis and therapy of human breast and ovarian cancer
PT and for diagnosing pre-disposition to these cancers
PS Claim 1; 218pp; English.
CC T17439-T17453 and T17455-T17529 represent mutations of the human breast
CC and ovarian cancer predisposing gene (BRCA1) (for wild type see T17438).
CC Proteins encoded by these mutations (see R81493-R81497 and R81499-R81546)
CC can be used as immunogens for antibody production. These mutant BRCA1
CC genes have at least 1 mutation or polymorphism in comparison to the wild
CC type sequence. By detecting a germline alteration in the wild type BRCA1
CC gene, a predisposition for breast and ovarian cancer can be diagnosed.
CC In one method, BRCA1 mRNA isolated from a tissue sample from a subject
CC has a probe, corresponding to a fragment of the wild type sequence (or an
CC allele-specific probe for one of these mutations), added to it. The
CC conditions allow for hybridisation of the probe to the mRNA, and any
CC hybridisation which occurs is detected. Alternatively the BRCA1 gene in
CC the tissue sample is isolated, and a shift in electrophoretic mobility of
CC single stranded DNA from the sample on a non-denaturing polyacrylamide
CC gel indicates a mutation. These methods of detection can also diagnose a
CC lesion neoplasia associated with the BRCA1 locus. The methods may be
CC used in gene therapy, protein replacement therapy and protein mimetics,
CC and may be used to screen for drugs in cancer therapy.
CC Sequence 5914 BP; 2007 A; 1156 C; 1315 G; 1436 T;

```

```

alignment_scores:
  Quality: 121.50      Length: 436
  Ratio: 0.620        Gaps: 19
  Percent Similarity: 44.954  Percent Identity: 19.037

alignment_block:
US-08-656-811A-1 x T17502 ..
Align seg 1/1 to: T17502 from: 1 to: 5914

13 AlaArgGluTrpGlyLeuGluMetProValValGlnThrAspGlyGlnPh 29
   ::::::::::: ::::::::::: |||||::: ::: ||::::: ||:::
2637 AGTCGGGAACAAGCATAGAAATGGGAAGAAAGTGAACCTTGATGCTCAGTA 2686
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
29 eGlyAspLeuLysSerThr..... 35
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
2687 T.....TTGCAGATACATTCAAGGTTTCAAAAGCCGACGATTTGGCTC 2730
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
36 .....SerArg 37
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
2731 CGTTTTCAATCCAGGAAATGCAGAGAGGAATGTCAACATTTCTCTGCC 2780
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
38 HisGlyGlyAspGluSerLeuGlnProGlnGlyAlaThrLeuLys 54
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
2781 CACTCTGGG.....TCCTTAAGAAACAAGTCCAAAGAGTCACCTTTTGA 2824
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
54 sLeuGluProPheGluGluAspValLeuGlyAlaGluTrpMetGluSerS 71
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
2825 ATGTGAACAAGGAAGAAAT..... 2846.
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
71 eAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeu 87
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
2847 .....CAAGGAAAGAATGAGTCTAATATC 2870
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
88 HisProPheGluSerAsnLeuLeuGluPheThrSerLeuIleThrProAs 104
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
2871 AAGCCTGTACAGACA..... 2885
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
104 pAspSerThrValSerLysAspIleLeuSerSerThrLeuGlnPhePro 121
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
2886 .....GTTAATATCACTGCAGGCTTTCCCTG 2910
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
121 hr.....GlnProValAsnIleProLeuTyAlaSerHis 132
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
2911 TGGTTGGTCAGAAAGATAAGCCAGTTGATAATGCCAAATGTAGTACAAA 2960
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
133 GlyAlaGluAspPheSerAlaGluThrGluPhe.....GluAsnH1 146
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
2961 GGAGGCTCTAGGTTTGTCTATCATCTCAGTTCAGAGGCAACGAAACTGG 3010
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
146 sLeuSerProProAsp.....SerProGluGlnValA 157
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
3011 ACTCATTACTCCAAATAAATACATGGAGCTTTTACAAAACCCATATCGTATAC 3060
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
157 lProValIleAsnLeuGluProValGluLeuThrAlaSerHisMetThr 173
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
3061 CACCACCTTTTCCCATCAAGTCATTGTTTAAAACTAAATCTAAGAAAAAT 3110
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
174 ValIle.....SerProAspGlyLeuLe 181
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
3111 CTGCTAGAGGAAACCTTTGAGGACACATTCAATGTCACCTGAAGAGAAAT 3160
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
181 uGlyGlyMetGluLeuAlaSerGluSerLeuThrPheThr..... 194
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
3161 GGGAAATGAGACATTCACCAAGTACAGTGCACACATTAGCCGTAATAACA 3210
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
195 .....GluLeuAspPheValAsnPheAsnAspSerAlaValGlySerIle 209
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
3211 TTAGACAAAAATGTTTTTAAAGAAGCAACTCAAGCAATATTATTAAAGTA 3260
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
210 GlyGlyAlaGluLeuLeuGlySerProLeu.....Se 221
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::

```



```

626 .....AGCGCGGCACCACTTGCCTCCCAACAGATCCC 556
273 exLysSerLysGlnSerValLysThrSerAspAlaLysAlaProArgLys 289
657 GGTGCAGC.....ACCGCGGCTGCAAGCCCTCAAGGAAG 691

290 ThrArgThrProAla.....GlnProValProG 299
692 AGCGCGAGACCTGCGGAGATGCGGGAGAGACGCCGCCCTGTCCCT 741
299 LuHisValIleMetGluHisLeuAsp.LysLysAspArgLysLysLeuG1 315
742 ATCGACATGGATCTCAGGAGCGGATCAAGGAGAGAGAGAGCGGATGAG 791
315 nAsnLysAsnAlaAlaIleArgTyrArgMetLysLysLysGlyGluAlaG 332
792 GAACCGCATTCGCCCTCCAAAGTCCGGAAGAAAGAGCTGGAGCGATCG 841
332 InGlyIleLysGlyGluGluGlnGluLeuGluGluLeuAsnThrLysLeu 348
842 CTCGGCTAGAGAAAGTGAACCTTGAAGCGCAAAACTCCGAGCTG 891
349 LysThrLysValAspLeuGlnArgGluIleLysTyrMetLysAsnLe 365
892 GCATCCAGCGCAACATGCTCAGGAACAGGTGGCACACCTTAAG..... 936
365 uMetGluAspValCysLysAla.....LysGlyIleGlnLeuLysMet 379
937 ....CAGAAAGTCATGAGGCATCCACGAGTGGTGGTCCCAACTCATGCTA 981

seq_name: N_Geneseq_36:Q13674

```

```

seq_documentation_block:
AC Q13674 standard; DNA; 3518 BP.
DE 28-NOV-1991 (first entry)
DE E2A.E12 transcript.
KW Immunoglobulin; enhancer-binding factor; chromosomal translocation;
KW leukemia; fusion protein; ss.
OS Homo sapiens.
FH Key . Location/Qualifiers
FT cds 31..1992 /*tag= a
FT FT /label= E2Aalpha
FT FT /note= "mol. wt. 67.6 kD"
FT poly_a_signal 2513..2518 /*tag= b
FT W09113172-A.
PD 05-SEP-1991.
PF 22-FEB-1991. U01168.
PR 23-FEB-1990; US-484063.
PA (STRD ) LELAND STANFORD JR UNIV.
PA (WHIT-) WHITEHEAD INST BIOMED RE.
PI Cleary ML, Wellentin JD, Baltimore D, Murre C, Mccaw P;
DR WPI; 91-281484/38.
DR P-PSDB; R13950.
PT Detection of t(1;19) break-point-associated genes E2A and pml -
PT in chromosomal translocation, and prods, useful in diagnosis and
PT therapy of human neoplasm, esp. acute lymphoblastic leukaemia
PS Disclosure; Fig 4A; 104pp; English.
CC The cDNA contains a single ORF encoding 654 amino acids and 2.4 kb
CC of 3' untranslated mRNA. The 3' untranslated region is not fully
CC represented due to discontinuous reproduction in the specification.
CC With the exception of a single in-frame insertion of 3 bases,
CC nucleotides 678-2015 are identical in sequence to the nucleotides
CC 9-1343 of the E12 isolate of the E12 gene (Q13670). The inserted
CC triplet results in the addition of glutamine 390.
CC See also Q13669-75.
SQ Sequence 3518 BP; 703 A; 1166 C; 1020 G; 629 T;

```

alignment_scores:

```

Quality: 120.50 Length: 411
Ratio: 0.600 Gaps: 13
Percent Similarity: 48.905 Percent Identity: 20.195

alignment_block:
US-08-656-811A-1 x Q13674 ..

Align seg 1/1 to: Q13674 from: 1 to: 3518

4 AspLeuTrpSerGluAspPheGln.....LeuAlaArgG1 15
706 GAGCTCTGAGTCCCGCGGCGGCGGCTTCGGGCCCATGCTGGGTGG 755
15 utrpgLysLeuGluMetProValGlnThrAspGlyGlnPheGlyAspL 32
756 GGGCTCATCCCGCTGCCCTCCCGCGGCTAGCGGCGGCTGGCGAGCA 805
32 euLysSerThrSerArgHisGlyGly.....AspGluSerLeuSer 45
806 GTGGAAAGCAGCAGCACGTTTGGTGGCTGCACACGACGAGCGTATGGGC 855
46 LeuGlnProGlnGlyAlaThrLeuLys..... 54
856 TACCAGCTCCATGGACGACAGGTGAACGGTGGGCTCCCATCTCCATCCTC 905
55 .....LeuGluProPheGluG 60
906 CTTCTCCTCAGCCCGCGGAGCCACGTACGGCGCGTCTCCAGGCCACACGC 955
60 luAspValLeuGlyAlaGluTrpMet.....GluSerSerAspLeu 73
956 CGCCTGTCAGCGGGCGGACACCTCCTGGGTCCCGAGGGACACAGCT 1005
74 GlySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeuHisProPh 90
1006 GGCAGCTCCGGGATGCCCTCGGCAAGACACTGGCCTCATCTACTACCCC 1055
90 eGluSerAsnLeuLeuGluPheThrSerLeuIleThrProAspAspSert 107
1056 GGATCACTCAAGCAATAACTTCTCGTCC.....AGCCCTTCTACCCCG 1099
107 hrValSerLysAspIleLeuSerSerThrLeuGlnPheProThrGlnPro 123
1100 TGGGCTCCCCCGAGGCGCTGGCAGGAACCTCACAGTGGCTCGAGCAGGA 1149
124 ValAsnIleProLeuTyrAlaSerHisGlyAlaGluAspPheSerAlaG1 140
1150 GCGCCCGGTCCCTTATCGCCCGACGTACGAGCGGGGTCTCCACGGCCTGCA 1199
140 uThrGluPheGluAsnHisLeuSerProProAspSerProGluGlnValA 157
1200 GAGTAAGATAGAAGACCACTG.....GACGAGGCCA 1231
157 laProValIleAsnLeuGluProValGluLeuThrAlaSerHisMetThr 173
1232 TCCAGTGTCTCCGACGCCACGCGCTGGGC...ACAGCGGCGGACATGCA 1278
174 ValIleSerPro.....AspGlyLeuLeuGlyGlyMe 184
1279 AGCTGTCTCCTGGCCACCGGCGCTGGGCTCAGGTTTCACCGGCCCAT 1328
184 tGluLeuAlaSerGluSerLeuThrPheThrGluLeuAspPheValAsnP 201
1329 GTCGTGGGTGGCGGCAC..... 1347
201 heAsnAspSerAlaValGlySerIleGlyGlyAla.....GluGluLeu 215
1348 .....GCAGGCGCTGGTGGAGGACGCCACCCCGGAGGAGCGGC 1383
216 LeuGlySerProLeuSerValAspValGluSerThrIleSerPheSe 232
1384 CTCGAGGCGGACCACTCATGTGCACAACACGCGGCGCTC..... 1425

```


2731 CGTTTCAATCCAGGAATGCAGAGAGGAATGTGCAACATTTCTCTGCC 2780
 38 HisGlyGlyAspGluSerLeuSerLeuGlnProGlnGlyAlaThrLeuLeu 54
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 2781 CACTCTGGG.....TCCTTAAGAAACAAAGTCCAAAGTCACTTTTGA 2824
 54 sLeuGluProPheGluGluAspValLeuGlyAlaGluTrpMetGluSerS 71
 : ||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 2825 ATGTCAACAAAAGGAAGAAAT..... 2846
 71 erAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeu 87
 |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 2847CAAGAAAGATGATGCTTAATATC 2870
 88 HisProPheGluSerAsnLeuGluPheThrSerLeuIleThrProAs 104
 |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 2871 AAGCCTGTACAGACA..... 2885
 104 pAspSerThrValSerLysAspIleLeuSerSerThrLeuGlnPheProT 121
 : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 2886GTTAATATCACTGCAGGCTTTCCTG 2910
 121 hr.....GlnProValAsnIleProLeuTyThrAlaSerHis 132
 : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 2911 TGGTGTGCAGAAAGATNAGCCAGTGTGATAATGCCAATGTATATCAAA 2960
 133 GlyAlaGluAspPheSerAlaGluThrGluPhe.....GluAsnHi 146
 |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 2961 GGAGGCTCTAGTTTGTCTCATCTCAGTTCAGAGCAACGAACTGG 3010
 146 sLeuSerProProAsp.....SerProGluGlnVala 157
 |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 3011 ACTCATTACTCCAATAAACATGGACTTTTACAAAACCCATATCGTATAC 3060
 157 laProValIleAsnLeuGluProValGluLeuThrAlaSerHisMetThr 173
 |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 3061 CACCACCTTTTCCCATCAAGTCATTTGTATAAACTAAATGTAAGAAAAAT 3110
 174 Val.....IleSerProAspGlyLeuLe 181
 : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 3111 CTGCTAGAGGAAACTTTGAGCAACATTCATATCACCTGAAGAGAAAT 3160
 181 uGlyGlyMetGluLeuAlaSerGluSerLeuThr..... 192
 : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 3161 GGGAAATGAGAAACATTCCAAGTACAGTACAGCACAAATTAGCCGTAATAACA 3210
 193PheThrGluLeuAspPheValAsnPheAsnAsp... 203
 |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 3211 TTAGAGAAAATGTTTTTAAAGAGCCAGCTCAAGCAATATTAAATGAAGTA 3260
 204 ...SerAlaValGlySerIleGlyAlaGluGluLeuLeuGlySerPr 219
 |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 3261 GGTTCAGTACTAATGAGTGGGTCCAGTATTAAATGAAATAGGTTC... 3308
 219 oLeuSerValAspValGluSerThrIleSerPheSer...GlyProS 235
 |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 3309AGTGATGAAACATTCAGCAAGCACTAGGTAGAAACAGAGGCCAA 3354
 235 er.....SerProGluThrSerGln 241
 : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 3355 AATTGAATGCTATCTAGATTAGGGGTTTTGCAACCTGAGGTATATAA 3404
 242 SerSerIleLeuGluSerSer.....ProGluLeuTyThrLys..... 253
 : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 3405 CAAAGTCTTCTCGAAGTAATTGTAGCAATCCTCAATATAAAAGCAAGA 3454
 254ValIleSerThrSerIleAspAlaSerLysArgPheS 267
 |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 3455 ATATGAAGAAGTACTTCAGACTGTTAATACAGAT.....TTCT 3492
 267 erProTyThrArgSerSerLysSerLysGlnSerValLysThrSerAsp 283
 |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 3493 TCCATATCTGATTTCAGATTAAGAACGCTATAGGAAGTAGTCAT 3542

284 AlaLysAlaProArgLysThrArgThrProAlaGlnProValProGluHi 300
 |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 3543 GCATCT.....CAGGTTTGTCTGAGACACCTGATGA 3574
 300 sValIleMetGluHisLeuAspLysLysLysAspArgLysLysLeuGlnAsn. 316
 : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 3575 CTTGTTAGATGTTGTTGAAATAAAGGAAGATACTACTGTTTGTGTAATG 3624
 317LysAsnAlaAlaIleArgTyArgMetLysLysLysGlyGlu 330
 : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 3625 ACATTAGGAAGTTCTGCTCTTTTAGCAAAAGCGTCCAGAAAGGAGAG 3674
 331AlaGlnGlyI 334
 |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 3675 CTTAGCAGAGCTCTAGCCCTTTCACCCATACACATTTGGCTCAGGGTTA 3724
 334 eLysGlyGluGlnGluLeuGluGluLeuAsnThrLysLysLysThrL 351
 : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 3725 CCGAAGAGGGGCCAAGAAATTAGAGTCTCTCAGAGAGAACTTATCTAGTG 3774
 351 ysValAspAspLeu 355
 : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 3775 AGGATGAAGAGCTT 3788
 seq_name: N_Geneseq_36:V32452
 seq_documentation_block:
 ID V32452 standard; cDNA; 8982 BP.
 AC V32452;
 DT 12-OCT-1998 (first entry)
 DE Human receptor tyrosine kinase LMR2_h cDNA.
 KW Receptor tyrosine kinase; LMR2_h; human; signal transduction;
 KW cancer; neurodegenerative disorder; gene therapy; ss.
 OS Homo sapiens.
 FH Key
 FT CDS
 FT sig_peptide
 FT mat_peptide
 FT WO9822507-A2;
 PN 28-MAY-1998;
 PD 21-NOV-1997; U22526.
 PE 22-NOV-1996; US-031675.
 PR (SUGB-) SUGEN INC.
 PA Joho KE, Plowman GD;
 PI WPI; 98-312419/27.
 DR P-PSDB; W48845.
 DR New isolated receptor tyrosine kinase genes - which are expressed in
 PT neuronal tissues and tumour cells, useful as targets for
 PT neurodegenerative disorders or cancers
 PS Clalm 1; Fig 6D-H; 87pp; English.
 CC This nucleic acid molecule (NAM) codes for LMR2_r (see W48845), a
 CC novel human receptor tyrosine kinase (RTK). A partial cDNA clone
 CC was isolated from human heart cDNA using primers (see V42006 and
 CC V42010) based on rat LMR1_r cDNA (see V32448). 2 LMR2_h cDNA
 CC clones of 4349 and 5482 bp were isolated from an NCI-H460 human
 CC lung carcinoma cell line cDNA library and from a SNB75 library,
 CC respectively, using this partial clone. These clones span the
 CC complete 8982 bp human LMR2 cDNA. NAMS (see V32448-56) coding for
 CC novel kinases LMR1, LMR2 and LMR3 (see W48841-49) have been obtained
 CC from rat, human and mouse sources. Expression of LMR1 and LMR3 is
 CC highly restricted to neuronal tissues with minimal expression in
 CC other adult or embryonic organs or in human tumour cell lines.
 CC LMR2 expression is limited to adult neuronal tissues, but is also
 CC very abundantly expressed in other non-neuronal foetal tissues and
 CC in numerous tumour cell lines. Based on restricted expression of
 CC all 3 LMRs to adult neuronal tissues and the up-regulation of LMR2
 CC in a wide variety of tumour cell lines, these proteins may be
 CC critical targets for neurodegenerative disorders or cancer. Claimed
 CC probes and antibodies may be used for detecting neurodegenerative


```

seq_documentation_block:
ID T70064 standard; cDNA; 5656 BP.
AC T70064;
DT 27-AUG-1997 (first entry)
DE Mutant BRCA1 allele #5803.
KW BRCA1; breast cancer gene; protein truncation; deletion; probe;
KW genetic marker; hybridisation assay; screening; susceptibility;
KW ovarian; prostatic; chromosome 17q; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT 120..200
FT cds /*tag= a
FT US5622829-A.
PD 22-APR-1997.
PF 08-DEC-1993; 163959.
PR 08-DEC-1993; US-163959.
PR 18-APR-1994; US-232535.
PR 20-OCT-1994; US-326983.
PR 19-APR-1995; US-425061.
PA (REGC) UNIV CALIFORNIA.
PI Friedman L, King M, Lynch E, Ostermeyer B;
PI Rowell S, Szabo C;
PI WPI: 97-244387/22.
DR P-PSDB; W10000.
DR BRCA1 allele(s) - useful as genetic markers for breast, ovarian and
PT prostatic cancers
PS Claim 1; Column 11-18; 63pp; English.
CC T70064-75 are BRCA1 mutant alleles from germline DNA and breast cancer
CC patient tumours from chromosome 17q-linked families. In the present
CC sequence the mutation has led to the deletion of exon 3 (54 bp) and the
CC generation of a stop codon corresponding to amino acid position 27 of
CC the wild type protein, leading to a predicted protein truncation (see
CC W10000). The nucleic acids can be used as probes in hybridisation assays
CC for screening patients for susceptibility to breast, ovarian or prostatic
CC cancer.
SQ Sequence 5656 BP; 1939 A; 1087 C; 1263 G; 1367 T;

alignment_scores:
Quality: 118.50 Length: 438
Ratio: 0.605 Caps: 20
Percent Similarity: 44.749 Percent Identity: 19.178

alignment_block:
US-08-656-811a-1 x T70064 ..
Align seg 1/1 to: T70064 from: 1 to: 5656

13 AlaArgGluTrpGlyLeuGluMetProValValGlnThrAspGlyGlnPh 29
:::||||| :::::||||| :::::|||||
2583 ACTCGGAAACAAGCATAGAAATGGAAGAAAGTGAACCTTGATGCTCAGTA 2632
29 eGlyAspLeuLysSerThr.....
: ||:::|||||
2633 T.....TTGCAGATACATTCAAGGTTTCAAGGCCCGCATGTTTGCTC 2676
36 .....
2677 CGTTTTCATCCGAATGCAGAGAGGAATGTCACATCTCTGCC 2726
38 HisGlyGlyAspGluSerLeuSerLeuGlnProGlnGlyAlaThrLeuLys 54
||||||| |||||
2727 CACTCTGG.....TCCTTAAGAAACAAGTCCCAAGATCATTTCATCA 2770
54 sLeuGluProPheGluGluAspValLeuGlyAlaGluTrpMetGluSers 71
: ||| |||||
2771 ATGTGAACAAAAGGAAGAAAT..... 2792
71 erAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeu 87
||| |||
2793 .....CAAGGAAGAATGAGTCTAATATC 2816

```

```

88 HisProPheGluSerAsnLeuLeuGluPheThrSerLeuLeuThrProAs 104
||| :::::
2817 AAGCCTGTACAGACA..... 2831
104 pAspSerThrValSerLysAspIleLeuSerSerThrLeuGlnPhePro 121
:::|||||
2832 .....GTTAATATCACTGCAGGCTTTCCTG 2856
121 hr.....GlnProValAsnIleProLeuTyrAlaSerHis 132
:::|||||||
2857 TGGTGGTCCAGAAAGATAAGCCAGTTGATAATGCCAAATGCTAGTATCAA 2906
133 GlyAlaGluAspPheSerAlaGluThrGluPhe.....GluAsnHI 146
||||||| |||
2907 GGAGGCTCTAGGTTTGTCTATCATCTCAGTTCAGAGGCAACGAACTGG 2956
146 sLeuSerProProAsp.....SerProGluGlnValA 157
||| |||||
2957 ACTCATTTACTCCAAATAAACAATGGAGCTTTTACAAAACCCCATATCGTATAC 3006
157 laProValIleAsnLeuGluProValGluLeuThrAlaSerHisMetThr 173
||||||| :::::
3007 CACCAGTTCCTCCATCAAGTCATTTGTTAAACTAAATGTAAGAAAAAT 3056
174 ValIle.....SerProAspGlyLeuLe 181
:::|
3057 CTGCTAGAGGAAAACTTTGAGGAACATTCAAATGTCACTGAAAGAGAAAT 3106
181 uGlyGlyMetGluLeuAlaSerGluSerLeuThr..... 192
||| |||
3107 GGGAAATGAGACATTTCCAAAGTACAGTGAGCACAAATTAGCCGTAAATAACA 3156
193 .....PheThrGluAspPheValAsnPheAsnAsp... 203
||| |||
3157 TTAGACAAATGTTTTTAAGAACGCCAGCTCAAGCAATATTATGAAGTA 3206.
204 ...SerAlaValGlySerIleGlyAlaGluGluLeuLeuGlySerPr 219
|||||
3207 GGTTCAGTACTAATGAAGTGGCTCCACTATTATTAAATAAGTTTCC... 3254
219 oLeuSerValAspValGluSerThrIleSerPheSer...GlyProS 235
|||
3255 ...AGTGATGAAACATTTCAAGCAGAACTAGGTAGAAACAGAGGCCAA 3300
235 er.....SerProGluThrSerGln 241
:::|
3301 AATTGAATGCTATGTTAGATTAGGGGTTTTCACACCTGAGGTCTATAAA 3350
242 SerSerIleIleGluSerSer.....ProGluLeuTyrLys.... 253
|||||
3351 CAAAGTCTTCTCGGAAGTAATTGTAAGCATCCTGGAATAAAAAAGCAAGA 3400
254 .....ValIleSerThrSerSerIleAspAlaSerLysArgPheS 267
|||||
3401 ATATGAAGAAGTAGTTTCAGACTGTTAATACAGAT.....TTCT 3438
267 exProThrSerArgSerLysSerLysGlnSerValLysThrSerAsp 283
|||||
3439 CTCCATATCTGATTTCAGATAACTTAGAACACGCCCTATGGGAAGTAGTCAT 3488
284 AlaLysAlaProArgLysThrArgThrProAlaGlnProValProGluHI 300
|||||
3489 GCATCT.....CAGGTTTGTCTGACACACCTCATGA 3520
300 sValIleMetGluHisLeuAspLysLysAspArgLysLysLeuGlnAsn.. 316
|||||
3521 CCTGTTAGATGATGGTGAATAAAGGAAGATACTAGTTTGTCTGAAAAATG 3570
317 .....LysAsnAlaAlaIleArgTyrArgMetLysLysLysGlyGlu 330
|||||
3571 ACATTAAGGAAAGTTCTGCTGTTTGTAGCAAAAGCCTCCAGAAAGAGAG 3620
331 .....AlaGlnGlyI 334

```

```

3621 CTTAGCAGGAGTCTAGCCCTTTCACCCATACACATTTCAGGGTTA 3670
|||||
334 eLysGlyGluGluGlnGluLeuGluLeuAsnThrLysLysThrL 351
::: |||||::: |||||::: |||||::: |||||::: |||||:::
3671 CCGAGAGGGCCCAAGAAATAGAGTCTCAGAGAGAACTATCTAGTG 3720
351 ysValAspAspLeu 355
::: |||||::: |||||::: |||||::: |||||::: |||||:::
3721 AGGATGAAGAGCTT 3734

seq_name: N_Geneseq_36.V60560

seq_documentation_block:
ID V60560 standard; cDNA; 5656 BP.
AC V60560;
DT 09-DEC-1998 (first entry)
DE Tumorigenic BRCAL allele #5803.
KW Tumorigenic allele; ovarian; breast cancer susceptibility gene; BRCAL;
KW identification; specific mutation; allele specific antibody; detection;
KW binding assay; treatment; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS
FT US821328-A.
PN PD 13-OCT-1998.
PF 02-APR-1997; 825886.
PR 08-DEC-1993; US-163959.
PR 18-APR-1994; US-232535.
PR 20-OCT-1994; US-326983.
PR 19-APR-1995; US-425061.
PR 02-APR-1997; US-825886.
PA (REGC ) UNIV CALIFORNIA.
PI Friedman L, King M, Lee M, Lynch E, Ostermeyer B,
PI Rowell S, Szabo C;
DR P-PSDB; W79886.
PT Poly(peptide(s) based on mutated BRCAL allelic sequences - useful
PT for identification of specific allelic mutation(s) of the gene
PT causing breast cancer
PS Example 2; Columns 11-18; 62pp; English.
CC The present sequence represents a tumorigenic allele of the breast and
CC ovarian cancer susceptibility gene (BRCAL). Exon 3 has been deleted
CC in the present allele, which results in a truncated protein. Several
CC other mutagenic alleles have also been identified (see V60560-71). The
CC polypeptides are useful for the identification of specific mutations
CC of BRCAL. The alleles are specific mutations of the C-terminal end of
CC BRCAL, and can be used to compare with the translation product from a
CC patient to identify the mutation that is causing the disease. In this
CC respect, they can also be used to raise allele specific antibodies.
CC They can also be used for detection purposes. The antibodies can be
CC used in binding assays e.g. ELISA. The polypeptides can also be used
CC for the treatment of breast cancer.
SQ Sequence 5656 BP; 1939 A; 1087 C; 1263 G; 1367 T;

alignment_scores:
Quality: 118.50 Length: 438
Ratio: 0.605 Gaps: 20
Percent Similarity: 44.749 Percent Identity: 19.178

alignment_block:
US-08-656-811A-1 x V60560 ..
Align seg 1/1 to: V60560 from: 1 to: 5656

13 AlaArgGluTrpGlyLeuGluMetProValValGlnThrAspGlyGlnPh 29
::: |||||::: |||||::: |||||::: |||||::: |||||:::
2583 AGTCGGGAACACAGCATAGAATGGAAGAAAGTGAAGTGTGTCAGTA 2632
29 eGlyAspLeuLysSerThr.....
::: |||||::: |||||::: |||||::: |||||::: |||||:::

```

```

2633 T.....TTGCAGAAATCAAGTTTCAAGCGCCAGTCATTGTGTC 2676
36 .....SerArg 37
2677 CGTTTCAAATCCAGGAATCCAGAAGAGGAATGCAACATTTCTCTGCC 2726
38 HisGlyGlyAspGlySerLeuSerLeuGlnProGlnGlyAlaThrLeuLy 54
|||||::: |||||::: |||||::: |||||::: |||||:::
2727 CACTCTGGG.....TCCTTAAGAAACAAAGTCCAAAGTCACCTTTTGA 2770
54 sLeuGluProPheGluGluAspValLeuGlyAlaGluTrpMetGluSerS 71
: ||| |||||:::
2771 ATGTGAACAAAAGGAAGAAAT..... 2792
71 erAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeu 87
2793 .....CAGGAAGAAATGAGTCTAATATC 2816
88 HisProPheGluSerAsnLeuGluPheThrSerLeuLeuThrProAs 104
|||::: |||||:::
2817 AAGCCTGTACAGACA..... 2831
104 pAspSerThrValSerLysAspIleLeuSerSerThrLeuGlnPheProT 121
::: |||||::: |||||:::
2832 .....GTTAATATCATCTGCAGGCTTCTCTG 2856
121 hr.....GlnProValAsnIleProLeuTyrAlaSerHis 132
::: |||||::: |||||:::
2857 TGGTTGGTCAGAAAGATAGCCAGTTGATAATGCCAAATGTAGTATCAAA 2906
133 GlyAlaGluAspPheSerAlaGluThrGluPhe.....GluAsnHi 146
|||||::: |||||::: |||||::: |||||:::
2907 GGAGCCTCTAGTTTGTCTATCATCTCAGTTCAGAGCAACAGAACTGG 2956
146 sLeuSerProProAsp.....SerProGluGlnVala 157
|||::: |||||::: |||||::: |||||:::
2957 ACTCATTTACTCCAATAAATGACATGTTTACAAACCCATATCGTATAC 3006
157 laProValIleAsnLeuGluProValGluLeuThrAlaSerHisMetThr 173
|||||::: |||||::: |||||::: |||||:::
3007 CACCATTCTTTCCCATCAAGTCATTTGTTAAAACTAAATGTAAGAAATA 3056
174 ValIle.....SerProAspGlyLeuLe 181
::: |||||::: |||||::: |||||:::
3057 CTGCTAGAGGAAACTTTGAGGAACATTCATGTCCACCTGAAAGAGAAAT 3106
181 uGlyGlyMetLeuLeuAlaSerGluSerLeuThr..... 192
|||||::: |||||::: |||||::: |||||:::
3107 GGAATGAGAAACATTCAGATACAGTACAGCACAATTAGCCGTAATAACA 3156
193 .....PheThrGluLeuAspPheValAsnPheAsnAsp... 203
|||||::: |||||::: |||||::: |||||:::
3157 TTAGAGAAATGTTTAAAGAACCCAGCTCAAGCAATATTAAATGAAGTA 3206
204 ..SerAlaValGlySerIleGlyAlaGluGluLeuLeuGlySerPr 219
|||||::: |||||::: |||||::: |||||:::
3207 GGTTCACAGTAAATGAAGTGGCTCCAGTATTAAATGAATAGGTTC... 3254
219 oLeuSerValAspAspValGluSerThrIleSerPheSer...GlyProS 235
|||::: |||||::: |||||::: |||||:::
3255 ...AGTCATGAAACATTCAGAGCAACTAGGTAGAAACACAGAGGGCCA 3300
235 er.....SerProGluThrSerGln 241
::: |||||::: |||||::: |||||::: |||||:::
3301 AATTGAATGCTATGCTTAGATTAGGGTTTGGCAACCTGAGGTCTATAAA 3350
242 SerSerIleLeuGluSerSer.....ProGluLeuTyrLys..... 253
::: |||||::: |||||::: |||||::: |||||:::
3351 CAAAGTCTCTCTGGAGTAATGTAAGCATCTCTGAATAAAAAAGCAAGA 3400
254 .....ValIleSerThrSerSerIleAspAlaSerLysArgPheS 267
|||||::: |||||::: |||||::: |||||:::
3401 ATATGAAGAAGTAGTTTCAGACTGTTAATACAGAT.....TTCT 3438

```


This Page Blank (uspto)

Align seg 1/1 to: US-09-005-970-1 from: 1 to: 3190

19 GluMetProValGlnThrAspGlyClnPheGlyAspLeuYsSerTh 35
|||::: ||| ||| :::::|||
325 GAGCAGACGGTCCTGCATCGGACCAGCGAAGTGACGAGCTG..... 366

35 rSerArgHisGlyAspGluSerLeuSerLeuGlnProGlnGlyAlar 52
|||::: ||| |||:::||| |::: |||
367 .TCAGACCCGGAGACTGAGGCCCTCATGTACCACACGCACTTCTCGG 415

52 hrLeuLysLeuGluProPheGluGluAspValLeuGlyAlaGluTrpMet 68
||| |||::: |||:::||| |||:::|||
416 AGCTC...CTAGACGAGTTTCCCAGAAGCTCTCTGGCT..... 450

69 GluSerSerAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisGl 85
::: |||:::|||
451CAGTCTCTGAGTGAC..... 465

85 uArgLeuHisProPheGluSerAsnLeuGluPheThrSerLeuIleTr 102
||||| |||
466CCTTCTCTCA..... 477

102 hrProaspAspSerThrValSerLysAspIleLeuSerSerThrLeuGln 118
::: |||:::||| |||:::|||
478GAGAAGCGAGTCAATGAGGTGGAGCATCT..... 510

119 PheProThrGlnProValAsnIleProLeuTyrrAlaSerHisGlyAlaGl 135
|||||:::|||::: |||:::|||
511 ..CCAACATCACACGG....CCTCTCATCCAGGCTGAACACGACTA 551

135 uAspPheSerAlaGluThrGluPheGluAsn.....HisLeuSerP 149
::::: ||| |||::: |||::: |||
552 CTCTCTGAGGAGCGCCCGGACTAGTCACCATTACCCATCGGGCTA 601

149 roProaspSer.....ProGlnVal..... 156
||||| |||:::|||
602 CCAGCGACAGCTTCAATGACGAGGAGGTGGAGAGTAANAATGTTACCTG 651

157AlaProValIleAsnLeuGluProValGluLeuTh 168
|||:::|||:::|||:::|||
652 TCTACAGAGTTTCCTTCAGCTACATCAAGAAGACCAATC..... 693

168 rAlaserHisMetThrValIleSerProAspGlyLeuLeuClyglyMetG 185
||| |||:::||| |||:::|||
694ACAGAGAGCGCCCCCGGGACTT..... 717

185 luLeuAlaSerGluSerLeuThrPheThrGluLeuAspPheValasnPhe 201
::: ||| ::|||:::|||:::|||
718 ..GPCCTCTGTCTCCTCATCTGACCATCACAGCCATTCC...ACTCCTTT 762

202 Asn.....AspSerAlaValcylserIleGlyGlyAl 212
::: ||| |||:::|||:::|||
763 GAAAAAGAAAGAGTCCCTCTCGATATATAATCTGGGGGGACTCCTCATG 812

212 aGluGluLeuGlySer...ProLeuSerValAspValcluseSert 228
::: |||:::|||:::|||:::|||
813 CCAGACCTTATTCTTAGATTTAGCTGTGGAGCCCCCAGAGTGGATCAGT 862

228 hrIleSerPheSerGlyProSerSerProGluThrSerGlnSerSerFile 244
::|||:::|||:::|||:::|||
863 TCTTAAACTTCTCC...CCGAAAAGACCCCTCCGTGGGATCACTGCCACTTA 909

245 IleGluSerSerProGluLeuTyrrLysValIleSerThrSerSerIleAs 261
::: |||:::|||:::|||:::|||
910 CCACCAAACACCCAGCTAGCTAGCAGACGAGTGACTCTGAGGCGACTTTGAG 959

261 pAlaserLysArgPheSerProTyrrSerArgSerSer..... 273
::: |||:::|||:::|||:::|||
960 CCCC AACCCGCTGCATCCCTTCAGCCCTGTCTCAGGCCCCACAGCCCTG 1009

274LysSerLysGlnSerValLysThrSerAsp...Ala 284
::: |||:::|||:::|||:::|||

```

1010 TCAGAGCCATGCCCCGGGGCCCTCTGCCTTGTGCCACATCTCCTCTCCCTC 105
285 LysAlaProArgLysThrArg..... 291
1060 ACAGCTCCACATAAGCTGCAGGATCGGGCCCTGGTCTGTGACAGAGA 1109
292 .....ThrProAlaGlnProValProGluHisValI 302
1110 GGAGAGAGAGGACCCCTGTTGCCGAGGCTATCCCATTCGCCACCAAGCTGC 1159
302 LeMet.....GluHisLeuAspLysLysAspArgLysLysLeu 314
1160 CTTCTGACAAATCTGAGAGAGAGCCCTCAAGAAAAATCCGGAGAAAAATC 1209
315 GlnAsnLysAsnAlaAlaIleArgTyrArgMetLysLysLysGlyGluAl 331
1210 AAGAATAAGATTCTCCCCAAGAAACGACGAGAGAAGAAAGAAAGAAATACAT 1259
331 aGlnGlyIleLysGlyGluGlnGluLeuGluLeuAsnThrLysL 348
1260 GGACAGCCTGGAGAAAAAGTGAGTCTTGTTCACACTGAGAACTGGAGC 1309
348 euLysThrLysValAspAspLeuGln.....Arg 357
1310 TTCGGAAGAGCTGGAGTCTGGAGAACCAATAGGACTCTCTTTCAG 1359
358 GluIleLysTyrMetLysAsnLeu.....MetGluAspValcy 370
1360 CAACCTTCAGAGCTTCAGACTTTGGTGTGGGAGAGGTCTCTCGAACCTG 1409
370 slsAlaLysGlyIleGln 376
1410 CAAGTAGCTGGCACACAG 1428

seq_name: /cgn2_6/ptodata/2/1na/5D_COMB.seq:US-08-319-866-7

seq_documentation_block:
; Sequence 7, Application US/08319866
; Patent No. 5929223
; GENERAL INFORMATION:
; APPLICANT: Tully, Timothy P.
; APPLICANT: Yin, Jerry C.
; TITLE OF INVENTION: REGULSKI, Michael
; TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF GENES
; TITLE OF INVENTION: ASSOCIATED WITH LONG-TERM MEMORY
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/319,866
; FILING DATE: 7-OCT-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: CSHL94-03
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 7:

```


;
STRANDEDNESS: double-stranded
TOPOLOGY: linear
US-08-993-1118-9

alignment_scores:
Quality: 124.50 Length: 372
Ratio: 0.688 Gaps: 17
Percent Similarity: 48.656 Percent Identity: 22.043

alignment_block:
US-08-656-811A-1 x US-08-993-1118-9 ..

Align seg 1/1 to: US-08-993-1118-9 from: 1 to: 4225

5 LeuTrpSerGluAspPheGlnLeuAlaArgGluTrpGlyLeuGluMetPr 21
||| :|||:||||| :|||: |||: |||: |||:
2102 CTCAGGGGGAGGAATCCAGTCCTCTCTCCAGAGCCC 2139

21 ovalValGlnThrAspGlyGlnPheGlyAspLeuLysSerThrSerArgH 38
||| :|||: |||: |||: |||: |||: |||: |||:
2140 TCTGAGCATCTGCTCC.....TCTCCACTCCATCCAGTCTTCCCC 2180

38 IsGlyGlyAspGluSerLeuSerLeuGlnProGlnGlyAlaThrLeuLys 54
:|||: |||: |||: |||: |||: |||: |||: |||:
2181 AGAGTTTCCCTGAGAGTCTCAGAGTCTCTGAGGGGCGCTGTC..... 2224

55 LeuGluProPheGluGluAspValLeuGlyAlaGluTrpMetGluSerSe 71
:|||: |||: |||: |||: |||: |||: |||: |||:
2225 CAGTCTCTCTCCATAGTCTCCTCAGAGCCCTCTGAGGGGATGCACTCCCA 2274

71 rAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeuH 88
:|||: |||: |||: |||: |||: |||: |||: |||:
2275 ATCTCTCTCCAGAGTCTCAGAGTCTCCTGAGGGGAGGATTCCTGT 2324

88 isProPheGluSerAsnLeuLeuGluPheThrSerLeuLeuThrProAsp 104
|||: |||: |||: |||: |||: |||: |||: |||:
2325 CTCCTCTCCAA.....ATTCTCTCAG 2344

105 AspSerThrValSerLysAspIleLeuSerSerThrLeuGlnPhePro.. 120
:|||: |||: |||: |||: |||: |||: |||: |||:
2345 AGTCTCTTGAGGGAGGAGACTCCTGCTCTCT...CTCCATTTCTCTCA 2391

121ThrGlnProValAsnIleProLeuT 129
:|||: |||: |||: |||: |||: |||: |||: |||:
2392 GAGTCTCTGAGTGGGAGGACTCCTCTCTCTCTCTCCACTTCTCTCAGT 2441

129 yrAlaSerHisGlyAlaGluAspPhe..... 137
:|||: |||: |||: |||: |||: |||: |||: |||:
2442 TTCTCTCTCAGGGG...GAGGACTTCCAGTCTTCTCTCCAGAGTCTCTGTG 2488

138SerAlaGluThrGluPhe.....G1 144
|||: |||: |||: |||: |||: |||: |||: |||:
2489 AGTATCT 2538

144 uAsnHisLeuSerProProAspSerProGluGlnValAlaProValIleA 161
:|||: |||: |||: |||: |||: |||: |||: |||:
2539 GAGTCTCAGAGTCTCTCTGAGGGCCCTGCTCAG...TCTCTCTCTC... 2581

161 snLeuGluProValGlu.....Leu 167
:|||: |||: |||: |||: |||: |||: |||: |||:
2582 ..CAGAGACCTGTCTCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2629

168 ThrAlaSerHisMetThrValIleSerProAspGlyLeuLeuGlyGlywe 184
:|||: |||: |||: |||: |||: |||: |||: |||:
2630 CAAAGTTCCCATGAGAGTCTCAGAGTCTCTCTCTCTCTCTCTCTCTCTCT 2665

184 tGluLeuAlaSerGluSerLeuThrPheThrGluLeuAspPheValAsnP 201
:|||: |||: |||: |||: |||: |||: |||: |||:
2665 2665

201 heAsnAspSerAlaValGlySerIleGlyAlaGluLeuLeuGly 217
||| |||: |||: |||: |||: |||: |||: |||: |||:
217

2666GGCCCTGCCAGTCTCTCTCTCCAG 2689
218 SerProLeuSer.....ValAspAspValGluSerThrIleSerPheSe 232
|||: |||: |||: |||: |||: |||: |||: |||:
2690 AGTCTCTGAGCTCTCTCCCTCCTCCACTTCATCGAGTCTTTCCAGAG 2739
232 rGlyPro.....SerSerProGluThrSerGlnSerSerIleIleGluS 247
:|||: |||: |||: |||: |||: |||: |||: |||:
2740 TTCTCTGTGAGCTCTCTCCCTCCTCCACTTCATCGAGTCTTTCCAGA 2789
247 erSerProGluLeuTyryLysValIleSerThrSerSerIleAspAlaSer 263
|||: |||: |||: |||: |||: |||: |||: |||:
2790 GTTCCCTGAGAGTCTCTCCAGAGTCTCTGTGATCTCTCTCTCTCTCC 2839
264 LysArgPheSerProTyzSerArgSer..... 273
:|||: |||: |||: |||: |||: |||: |||: |||:
2840 ACTTCTGAGCCCATTCAGTGAAGAGTCCAGCAGCCAGTAGATCAATA 2889
274LysSerLysGlnSerValLysThrSerAspAlaLysAlaProArgL 289
:|||: |||: |||: |||: |||: |||: |||: |||:
2890 TACAAGTTCTCAGACACTTCTAGAGAGTGTATCTTGTACACAGACGCG 2939
299 ysThrArgThrProAlaGlnProValProGluHisValIleMetGluHis 305
:|||: |||: |||: |||: |||: |||: |||: |||:
2940 ACTCTCTGATAGAGAGCGCCCTTGTTCACCTATACACTGGATGAAAG 2989
306 LeuAspLysLysAspArgLysLeuGlnAsnLysAsnAlaAlaIleArg 322
:|||: |||: |||: |||: |||: |||: |||: |||:
2990 GTGGAGAGTGTGGCGGTTCTTCTC.....CTCAA 3021
322 gTyrArgMetLysLys 327
:|||: |||: |||: |||: |||: |||: |||: |||:
3022 ATATCAAGTGAAGCAG 3037

seq_name: /cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-845-528C-9

seq_documentation_block:

; Sequence 9, Application US/08845528C
; Patent No. 6027924

; GENERAL INFORMATION:

; APPLICANT: LUCAS, Sophie;

; APPLICANT: DE SMET, Charles;

; APPLICANT: BOON-FALLEUR, Thierry

; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR

; TITLE OF INVENTION: REJECTION ANTIGEN PRECURSOR MAGE-C1 AND USES

; TITLE OF INVENTION: THEREOF

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Felfe & Lynch

; STREET: 805 Third Avenue

; CITY: New York City

; STATE: New York

; COUNTRY: USA

; ZIP: 10022

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage

; COMPUTER: IBM PS/2

; OPERATING SYSTEM: PC-DOS

; SOFTWARE: Wordperfect

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/845,528C

; FILING DATE: April 25, 1997

; CLASSIFICATION: 4335

; ATTORNEY/AGENT INFORMATION:

; NAME: Mary Anne Schofield

; REGISTRATION NUMBER: 36,669

; REFERENCE/DOCKET NUMBER: LUD 5455

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 688-9200

; TELEFAX: (212) 838-3884

; INFORMATION FOR SEQ ID NO: 9:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4225 base pairs


```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4249 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-480-784-21

alignment_scores:
  Quality: 118.50      Length: 438
  Ratio: 0.605        Gaps: 20
Percent Similarity: 44.749 Percent Identity: 19.178

alignment_block:
US-08-656-811A-1 x US-08-480-784-21 ..

Align seg 1/1, to: US-08-480-784-21 from: 1 to: 4249

13 AlaArgGluTrpGlyLeuGluMetProValValGlnThrAspGlyGlnPh 29
   :::::||||| :::::|||||
2048 AGTCGGAAACAAGCATAGAAATGGAAGAAAGTGAACCTTGATGCTCAGTA 2097
29 eGlyAspLeuLysSerThr..... 35
: |||:|||||
2098 T.....TTGCAGATAACATTCACGGTTTCAAAGGCCAGTCATTGCTC 2141
36 ..... 37
2142 CGTTTTCAAATCCAGAAATGCAGAGAGGAATGCAACATTCCTGCC 2191
38 HisGlyGlyAspGluSerLeuSerLeuGlnProGlnGlyAlaThrLeuLys 54
|||:||||| |||:|||||
2192 CACTCTGGG.....TCCTTAAAGAAACAAGTCCAAAGATCATTTCGA 2235
54 sLeuGluProPheGluGluAspValLeuGlyAlaGluTrpMetGluSerS 71
: ||| |||:|||||
2236 ATGTCGAACAAAAGGAAGAAAT..... 2257
71 eAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeu 87
||| |||:|||||
2258 .....CAAGGAAGAATGAGTCTAATATC 2281
88 HisProPheGluSerAsnLeuLeuGluPheThrSerLeuIleThrProAs 104
||| :::::
2282 AAGCCTGTACAGACA..... 2296
104 pAspSerThrValSerLysAspIleLeuSerSerThrLeuGlnPheProt 121
: ::::: ||| |||:|||||

```

```

2297 .....GTTAATATCACTGCAGGGCTTTCCTCG 2321
121 hr.....GlnProValAsnIleProLeuTyrAlaSerHis 132
   :::::|||||
2322 TGGTTGGTCAGAAAGATAAGCCAGTTGATAATGCCAAATGTAGTATCAAA 2371
133 GlyAlaGluAspPheSerAlaGluThrGluPhe.....GluAsnHl 146
|||:||||| |||:|||||
2372 GGAGGCTCTAGGTTTGTCTATCATCTCAGTTCAGAGGCACGNAAGTGG 2421
146 sLeuSerProProAsp.....SerProGluGlnVala 157
||| |||:|||||
2422 ACTATTACTCCAATAAATACATGGACTTTTACAAAACCCATATCGTATAC 2471
157 laProValIleAsnLeuGluProValGluLeuThrAlaSerHisMetThr 173
|||:||||| |||:|||||
2472 CACCACCTTTTCCCATCAAGTCATTGTTTAAAACTAAATGTAAGAAAAAT 2521
174 ValIle.....SerProAspGlyLeuLe 181
: |||:|||||
2522 CTGCTAGAGGAAAACTTTGAGGAACATTCAAATGTCACCTGAAGAGAAAT 2571
181 uGlyGlyMetGluLeuAlaSerGluSerLeuThr..... 192
|||:||||| |||
2572 GGAATGAGAAATTCCAAGTACACTGAGCACAAATTACCCGTAATAACA 2621
193 .....PheThrGluLeuAspPheValAsnPheAsnAsp... 203
||| ||| |||:|||||
2622 TTAGAGAAAATGTTTTTAAAGAGCCAGCTCAAGCAATATTATGAAGTA 2671
204 ...SerAlaValGlySerIleGlyAlaGluGluLeuGlySerPr 219
|||:||||| |||:|||||
2672 GGTTCAGTACTAATGAAGTGGCTCCAGTATTAATGAATAGGTTCC... 2719
219 oLeuSerValAspAspValGluSerThrIleSerPheSer...GlyProS 235
||| |||:|||||
2720 ....AGTGATGAAACATTCACAGCAGACTAGGTAGAAACAGAGGCCCAA 2765
235 ef.....SerProGluThrSerGln 241
: |||:|||||
2766 AATTGAATGCTATGCTTAGATTAGGGGTTTGCACACCTGAGGCTCTATAA 2815
242 SerSerIleGluSerSer.....ProGluLeuTyrLys..... 253
: |||:||||| |||:|||||
2816 CAAAGCTTCCTCGAAGTAATTGTAAGCATCCTGAAATAAAAAAGCAAGA 2865
254 .....ValIleSerThrSerSerIleAspAlaSerLysArgPheS 267
|||:||||| |||
2866 ATATGAAGAAGTAGTTACAGACTGTTAATACAGAT.....TTCT 2903
267 eProTyrSerArgSerLysSerLysGlnSerValLysThrSerAsp 283
|||:||||| |||:|||||
2904 CTCCTATCTGATTTCAGATACTTAGAACAGCCTATGGGAAGTAGTCAT 2953
284 AlaLysAlaProArgLysThrArgThrProAlaGlnProValProGluHl 300
|||:||||| |||:|||||
2954 GCATCT.....CAGGTTTGTCTGAGACACCTGATGA 2985
300 sValIleMetGluHisLeuAspLysLysAspArgLysLysLeuGlnAsn.. 316
|||:||||| |||:|||||
2986 CTTGTAGATGATGGTGAATAAAGGAAGATACTAGTTTTTGTGTAAGATG 3035
317 .....LysAsnAlaAlaIleArgTyrArgMetLysLysLysGlyGlu 330
|||:||||| |||:|||||
3036 ACATTAAAGAAAGTCTCTGTTTTTAGCAAAAGCGCTCCAGAAAGGAGAG 3085
331 .....AlaGlnGlyIl 334
|||:|||||
3086 CTAGCAGAGTCTTAGCCCTTTCACCCATACACATTTGGCTCAGGGTTA 3135
334 eLysGlyGluGlnGluLeuGluLeuAsnThrLysLysLeuLysThrL 351
: |||:||||| |||:|||||
3136 CCGAAGAGGGGCCAAGAAATTAGAGTCTCAGAAAGAACTTATCTAGTG 3185

```

```
351 ysValAspLeu 355
:: : : : :
3186 AGGATGAAGAGCTT 3199

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-483-553-21
seq_documentation_block:
; Sequence 21, Application US/08483553
; Patent No. 5709999
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,553
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4249 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
```

```
US-08-483-553-21

alignment_scores:
  Quality: 118.50      Length: 438
  Ratio: 0.605        Gaps: 20
  Percent Similarity: 44.749  Percent Identity: 19.178

alignment_block:
US-08-656-811A-1 x US-08-483-553-21

Align seg 1/1 to: US-08-483-553-21 from: 1 to: 4249

13 AlaArgGluTrpGlyLeuMetProValValGlnThrAspGlyGlnPh 29
:::||||| : : : : : : : : : : : : : : : : : : : : :
2048 AGTCGGGAACAAGCATAGAAATGCAACAAGTGAACCTTGATGCTCAGTA 2097
:::||||| : : : : : : : : : : : : : : : : : : : : :
29 eGlyAspLeuLysSerThr..... 35
: |||: : : : : : : : : : : : : : : : : : : : :
2098 T.....TTGCAGATACATTCAAGTTTCAAAGCGCCAGTCATTGTGCTC 2141
36 .....SerArg 37
|||
2142 CGTTTCAAATCCAGGAATGCAGAAGAGGAATGTGCAACATTTCTCTGCC 2191
|||
38 HisGlyGlyAspGluSerLeuGlnProGlnGlyAlaThrLeuLys 54
|||||: : : : : : : : : : : : : : : : : : : : :
2192 CACCTCTGGG.....TCCTTAAAGAACAAAGTCCAAAAGTCACCTTTGA 2235
54 sLeuGluProPheGluGluAspValLeuGlyAlaGluTrpMetGluSers 71
: ||| : : : : : : : : : : : : : : : : : : : : :
2236 ATGTGAACAAAAAGGAAGAAAT..... 2257
71 erAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeu 87
||| : : : : : : : : : : : : : : : : : : : : :
2258 .....CAAGGAAGAATGAGTCTAATATC 2281
88 HisProPheGluSerAsnLeuGluPheThrSerLeuLeuThrProAs 104
||| : : : : :
2282 AAGCCTGTACAGACA..... 2296
104 pAspSerThrValSerLysAspIleLeuSerSerThrLeuGlnPheProt 121
: : : : : : : : : : : : : : : : : : : : :
2297 .....GTTAATATCATCATCAGGCTTTCCTG 2321
121 hr.....GlnProValAsnIleProLeuTyrAlaSerHis 132
: : : : : : : : : : : : : : : : : : : : :
2322 TGGTGGTCAGAAAGATAGCCAGTTGATATGCCAAATGTAGTATCAAA 2371
133 GlyAlaGluAspPheSerAlaGluThrGluPhe.....GluAsnHI 146
|||: : : : : : : : : : : : : : : : : : : : :
2372 GGAGGCTCTAGGTTTGTCTATCATCTCAGTCAGAGCAACGAACTGG 2421
146 sLeuSerProProAsp.....SerProGluGlnValA 157
||| : : : : : : : : : : : : : : : : : : : : :
2422 ACTCATTACTCCAAATAACATGGACTTTTACAAAACCCATATCGTATAC 2471
157 laProValIleAsnLeuGluProValGluLeuThrAlaSerHisMetThr 173
|||: : : : : : : : : : : : : : : : : : : : :
2472 CACCACCTTTTCCCATCAAGTCATTTGTTAAAACTAAATGTAAAGAAAAT 2521
174 ValIle.....SerProAspGlyLeuLe 181
: : : : : : : : : : : : : : : : : : : : :
2522 CTGCTAGAGAGAAACTTTGAGGAACATTCATGTCCACCTGAAGAGAAAT 2571
181 uGlyGlyMetGluLeuAlaSerGluSerLeuThr..... 192
|||: : : : : : : : : : : : : : : : : : : : :
2572 GGGAAATGAGAACATTCACAGTACAGTACAGCAATTAGCCGTAATAACA 2621
193 .....PhethrGluLeuAspPheValAsnPhAsnAsp... 203
||| : : : : : : : : : : : : : : : : : : : : :
2622 TTAGAGAAAATGTTTTTAAAGAACGCCCTCAAGCAATATTAAATGAAGTA 2671
```

```
204 ....SerAlaValGlySerIleGlyGlyAlaGluGluLeuLeuGlySerPr 219
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2672 GGTTCAGTAAAGTAAAGTGGCCAGTAAATGAATAGTATGCTCC... 2719
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
219 oLeuSerValAspAspValGluSerThrIleSerPheSer...GlyProS 235
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2720 ....AGTGATGAACATTCAGACAGACTAGGTAGAACACAGAGGCCAA 2765
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
235 er.....SerProGluThrSerGln 241
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2766 AATTGAATGCTATGCTTAGATTAGGGTTTTCGAACCTGAGGTCTATAA 2815
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
242 SerSerIleGluSerSer.....ProGluLeuTyLys..... 253
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2816 CAAAGTCTTCCTCGGAAGTAATTTGTAAGCATCCTGAAATAAAAAAGCA 2865
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
254 .....ValIleSerThrSerSerIleAspAlaSerLysArgPheS 267
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2866 ATATGAAGAAGTAGTTCAGACTGTTAATACAGT .....TTCT 2903
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
267 erProTySerArgSerSerLysSerLysGlnSerValIysThrSerAsp 283
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2904 CTCCATATCTGATTTCAGATACTTAGACAGCCTATGGGAAGTAGTCAT 2953
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
284 AlaLysAlaProArgLysThrArgThrProAlaGlnProValProGluHi 300
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2954 GCATCT.....CAGGTTTGTCTGAGACACCTGATGA 2985
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
300 sValIleMetGluHisLeuAspLysLysAspArgLysLysLeuGlnAsn. 316
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2986 CCTGTAGATGATGTGAAATAAAGCAGATACTAGTTTGTCTGAAATG 3035
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
317 .....LysAsnAlaAlaIleArgTyArgMetLysLysLysGlyGlu 330
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
3036 ACATTAAGGAAGTCTGCTGTTTTAGCAAAAGCGTCCAGAAAGAGAG 3085
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
331 .....AlaGlnGlyI 334
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
3086 CTTAGCAGAGCTCTAGCCCTTCCACCATACACATTTGGCTCAGGGTTA 3135
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
334 eLysGlyGluGluGlnGluLeuGluGluLeuAsnThrLysLysLysThrL 351
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
3136 CCGAAGAGGGCCAGAAATAGACTCTTCAGAGAGAACTTATCTAGTG 3185
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
351 ysValAspAspLeu 355
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
3186 AGGATGAAGAGCTT 3199
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-487-002-21

seq_documentation_block:
; Sequence 21, Application US/08487002
; Patent No. 5710001
; GENERAL INFORMATION:
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Simard, Jacques
; APPLICANT: Emi, Mitsuru
; APPLICANT: Nakamura, Yusuke
; APPLICANT: Durocher, Francine
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487.002
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 4249 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORGANISM: Homo sapiens
US-08-487-002-21

alignment_scores:
Quality: 118.50 Length: 438
Ratio: 0.605 Gaps: 20
Percent Similarity: 44.749 Percent Identity: 19.178

alignment_block:
US-08-656-811A-1 x US-08-487-002-21 ..

Align seg 1/1 to: US-08-487-002-21 from: 1 to: 4249

13 AlaArgGluTrpGlyLeuGluMetProValValGlnThrAspGlyClnPh 29
:|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2048 ACTCGGAAACAGCATAGAAATGGAAGAAAGTGAACCTTGATGCTCAGTA 2097
:|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
29 eGlyAspLeuLysSerThr..... 35
:|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2098 T.....TTGCAGATACTTCAGGTTTCAAGGTTTCAAGGCCAGTCATTGCTC 2141
:|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
36 .....SerArg 37
:|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2142 CGTTTTCAATCCAGGAATGCAAGAGGAATGTCAACATTTCTCTGCC 2191
:|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
38 HisGlyGlyAspGluSerLeuSerLeuGlnProGlnGlyAlaThrLeuLy 54
:|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2192 CACTCTGGG.....TCCTTAAGAAACAAAGTCCAAAAGTCACTTTTGA 2235
:|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
54 sLeuGluProPheGluGluAspValLeuGlyAlaGluTrpMetGluSerS 71
:|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2236 ATGTGACACAAAGGAGAGAAAT..... 2257
:|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
71 erAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeu 87
:|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
```

```
2258 .....CAAGAAACAATGAGTCTAATATC 2281
88 HisProPheGluSerAsnLeuGluPheThrSerLeuIleThrProAs 104
2282 ANGCTGTACAGACA..... 2296
104 pAspSerThrValSerLysAspIleLeuSerSerThrLeuGlnPheProT 121
2297 .....GTTAATATCACTGCAGGCTTTCCTG 2321
121 hr.....GlnProValAsnIleProLeuThrAlaSerHis 132
2322 TGGTTGGTCAGAAAGATAGCCAGTTGTAATGCCAAATGTATATCAA 2371
133 GlyAlaGluAspPheSerAlaGluThrGluPhe.....GluAsnHi 146
2372 GGAGGCTCTAGTTTGTCTATCATCTCAGTTTCAGAGCAACGAACCTGG 2421
146 sLeuSerProProAsp.....SerProGluGlnValA 157
2422 ACTCATTTACTCCAAATAAACATGGACTTTTACAAAACCCATATCGTATAC 2471
157 laProValIleAsnLeuGluProValGluLeuThrAlaSerHisMetThr 173
2472 CACCACCTTTTCCCATCAAGTCATTGTGTAAAACTAAATGTAAAGAAAAT 2521
174 ValIle.....SerProAspGlyLeuLe 181
2522 CTGCTAGAGGAAACCTTTCAGGACATTCATGTCCCTGAAAGAAAT 2571
181 uGlyGlyMetGluLeuAlaSerGluSerLeuThr..... 192
2572 GGGAAATGAGAACATTCCAAGTACAGTGAGGCACCAATTAGCCGTAATAACA 2621
193 .....PheThrGluLeuAspPheValAsnPheAsnAsp... 203
2622 TTAGAGAAATGTTTAAAGAACGACGCTCAAGCAATATTAAATGAAGTA 2671
204 ...SerAlaValGlySerIleGlyAlaGluGluLeuGlySerPr 219
2672 GTTCCAGTACTAATGAGTGGCTCCAGTATTATGAATAGGTTC... 2719
219 oLeuSerValAspAspValGluSerThrIleSerPheSer...GlyProS 235
2720 ...AGTGATGAACAATTCAGACGACACTAGGTAGAAAACAGAGGCCAA 2765
235 er.....SerProGluThrSerGln 241
2766 AATTCAAATGCTATGCTTAGATTAGGGTTTTCGCAACCTGAGGTCTATAAA 2815
242 SerSerIleIleGluSerSer.....ProGluLeuTyrLys.... 253
2816 CAAAGTCTCTCGGAAGTAATGTAGGATCTCTGAATTAAMAAACAGA 2865
254 .....ValIleSerThrSerSerIleAspAlaSerLysArgPheS 267
2866 ATATCAAGAAGTAGTTCAGACTGTTTAATACAGAT.....TTCT 2903
267 erProTyrSerArgSerLysSerLysGlnSerValLysThrSerAsp 283
2904 CTCCATATCTGATTTCAGATACTTAGAACACCCCTATGGGAAGTAGTCAT 2953
284 AlaLysAlaProArgLysThrArgThrProAlaGlnProValProGluH 300
2954 GCATCT.....CAGGTTGTCTGAGACACCTGTATGA 2985
300 sValIleMetGluHisLeuAspLysLysAspArgLysLysLeuGlnAsn. 316
2986 CTGTTAGATGATGTGTAATAAGGAAGATACTAGTTTGTGTAAGAAATG 3035
317 .....LysAsnAlaAlaIleArgTyrArgMetLysLysLysGlyGlu 330
3036 ACATTAAAGAAAGTTCTGCTGTTTTCAGCAAAAGCGTCCAGAAAGGAGAG 3085
```

```
331 .....AlaGlnGlyIle 334
3086 CTTACAGAGAGTCTAGCCCTTTCACCCATACACATTTGGCTCAGGGTTA 3135
334 eLysGlyGluGluGlnGluLeuGluGluLeuAsnThrLysLeuLysThrL 351
3136 CGGAAGAGGGCCAAAGAAATTAGAGTCTCAGACAGAACTTATCTAGTG 3185
351 ysValAspAspLeu 355
3186 AGGATCAAGAGCTT 3199
seq_name: /cgn2_6/ptodata/2/ina/5b_COMB.seq:US-08-483-554B-21
seq_documentation_block:
; Sequence 21, Application US/08483554B
; Patent No. 5747282
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,554B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4249 base pairs
```

```
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-483-554B-21

alignment_scores:
  Quality: 118.50      Length: 438
  Ratio: 0.605        Gaps: 20
Percent Similarity: 44.749 Percent Identity: 19.178

alignment_block:
US-08-656-811A-1 x US-08-483-554B-21 ..
Align seg 1/1 to: US-08-483-554B-21 from: 1 to: 4249

13 AlaArgGluTrpGlyLeuGluMetProValValGlnThrAspGlyGlnPh 29
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2048 AGTCGGGAAACAGCATAGAAATGGAAGAAAGTGAACCTTGATGCTCAGTA 2097
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
29 eGlyAspLeuLysSerThr..... 35
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2098 T.....TTGCAGATACATTCAGGTTTCAAGGTTTCAAGCGCCAGTATTGGTCTC 2141
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
36 .....SerArg 37
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2142 CGTTTTCATCAGGAATGCAAGAGAGAAATGTCACATCTCTGCC 2191
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
38 HisGlyGlyAspGluSerLeuSerLeuGlnProGlnGlyAlaThrLeuLy 54
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2192 CACTCTGGG.....TCCTTAAAGAAACAAAGTCCAAAGTCACTTTTGA 2235
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
54 sLeuGlyProPheGluGluAspValLeuGlyAlaGluTrpMetGluSerS 71
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2236 ATGTGAACAAAGGAAGAAAT..... 2257
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
71 erAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeu 87
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2258 .....CAAGGAAGAAATGAGTCTAATATC 2281
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
88 HisProPheGluSerAsnLeuLeuGluPheThrSerLeuIleThrProAs 104
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2282 AAGCCTGTACAGACA..... 2296
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
104 pAspSerThrValSerLysAspIleLeuSerSerThrLeuGlnPheProT 121
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2297 .....GTTAATATCACTGCAGGCTTTCCTG 2321
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
121 hr.....GlnProValAsnIleProLeuTyrrAlaSerHis 132
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2322 TGGTGGTCAGAAAGATAGCCAGTTGTAATGCCAAATGTAGTATCAAA 2371
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
133 GlyAlaGluAspPheSerAlaGluThrGluPhe.....GluAsnHi 146
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2372 GGAGGCTCTAGGTTTGTCTATCATCTCAGTTCAGAGGCAACGAACTGG 2421
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
146 sLeuSerProProAsp.....SerProGluGlnVala 157
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2422 ACTCATTACTCCAAATAAATACATGGAGCTTTTACAAACCCATATCGTATAC 2471
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
157 laProValIleAsnLeuGluProValGluLeuThrAlaSerHisMetThr 173
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2472 CACCACCTTTTCCCATCAAGTCACTTTGTTAAAACTAAATGTAGAAAAAT 2521
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
174 ValIle.....SerProAspGlyLeuLe 181
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2522 CTGCTAGAGAAAACCTTTGAGGNACATTCATGCTCACCTGAAGAGAAAT 2571
```

```
181 uGlyGlyMetGluLeuAlaSerGluSerLeuThr..... 192
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2572 GGGAAATGAGAACATTCCAAGTACAGTGAGCACAATTAGCCGTAATAACA 2621
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
193 .....PheThrGluLeuAspPheValAsnPheAsnAsp..... 203
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2622 TTAGAGAAAATGTTTTTAAAGAAGCCAGCTCAAGCAGCATATTAATGAGTA 2671
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
204 ...SerAlaValGlySerIleGlyAlaGluGluLeuLeuGlySerPr 219
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2672 GGTTCCAGTACTAATGAAGTGGCTCCAGTATTAAATCAAAATAGTTTCC... 2719
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
219 oLeuSerValAspAspValGluSerThrIleSerPheSer...GlyProS 235
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2720 ...ACTGATGAAACATTCACAGCAAACTAGGTAGAAACAGAGGCCAA 2765
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
235 er.....SerProGluThrSerGln 241
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2766 AATTGAATGCTATGCTTAGATTAGGGGTTTTGCAACCTGAGGTCTATATAA 2815
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
242 SerSerIleIleGluSerSer.....ProGluLeuTyrrLys..... 253
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2816 CAAAGTCTCCTCGAAGTAATTAAGCATCTCTGAAATAAAAGCAAGA 2865
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
254 .....ValIleSerThrSerSerIleAspAlaSerLysArgPheS 267
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2866 ATATGAAGAAGTAGTTTCAGACTGTTAATACAGAT.....TTCT 2903
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
267 erProTyrrSerArgSerSerLysSerLysGlnSerValLysThrSerAsp 283
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2904 CTCATATCTGATTTCAGATAACTTAGACAGCGCTATGGGAAGTAGTCAT 2953
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
284 AlaLysAlaProArgLysThrArgThrProAlaGlnProValProGluH 300
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2954 GCATCT.....CAGGTTTGTCTGACACACCTGATGA 2985
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
300 sValIleMetGluHisLeuAspLysLysAspArgLysLysLeuGlnAsn. 316
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2986 CCTGTTAGATGATGGTCAATAAAGGAAGACTAGTGTTCCTGAAATG 3035
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
317 .....LysAsnAlaAlaIleArgTyrrArgMetLysLysGlyGlu 330
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3036 ACATTAGGAAGATCTGCTGTTTTTAGCAAAAGCGTCCAGAAAGGAGAG 3085
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
331 .....AlaGlnGlyI 334
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3086 CTTAGCAGGAGTCTAGCCCTTTCACCCATACACATTTGCTCAGGGTTA 3135
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
334 eLysGlyGluGluGlnGluLeuGluLeuAsnThrLysLysLeuThrL 351
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3136 CCGAAGAGGGGCCAAGAAATTAGAGTCTCTCAGAAGAGAACTTATCTAGTG 3185
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
351 ysValAspAspLeu 355
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3186 AGGATGAAGAGCTT 3199
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-488-011B-21

seq_documentation_block:
; Sequence 21, Application US/08488011B
; Patent No. 5753441
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
```


2904 CTCATATCTGATTTCAGATACTTAGACACGCTATGGGAAGTAGTCAT 2953
284 AlaAlaProArgLysThrArgThrProAlaGlnProValProGluHi 300
2954 GATCTCT.....CAGGTTTCTCTGAGACACCTGATGA 2985
300 sValileMetGluHisLeuAspLysLysAspArgLysLysLeuGlnAsn. 316
2986 CCGTGTAGTATGATGCTGAAATAGGAAGATAGTCTTGTGCTGAAATG 3035
317LysAsnAlaAlaLeuArgTyrArgMetLysLysLysGlyGlu 330
3036 ACATTAAGGAAGTCTGCTGTTTTAGCAAAAGCGTCCAGAAAGGAGAG 3085
331
3086 CTTAGCAGGAGTCTTACGCCCTTCCACCATACACATTTGGCTCAGGGTTA 3135
334 eLysGlyGluGlnGlnGluLeuGlnGluLeuAsnThrLysLeuLysThrL 351
3136 CCGAAGAGGGCCAGAAATAGTAGTCTCCTCAGAGAGAACTTATCTAGTG 3185
351 ysValAspAspLeu 355
3186 AGGATGAAGAGCTT 3199

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-10202-21

seq documentation_block:
Sequence 21. Application: PC/TUS9510202
GENERAL INFORMATION:
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Simard, Jacques
APPLICANT: Emi, Mitsuru
APPLICANT: Nakamura, Yusuke
APPLICANT: Dutcher, Francine
TITLE OF INVENTION: In vivo Mutations and Polymorphisms
TITLE OF INVENTION: In the 17q-linked Breast and Ovarian Cancer
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10202
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08-308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221

FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 4249 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
PCT-US95-10202-21

alignment_scores:
Quality: 118.50 Length: 438
Ratio: 0.605 Gaps: 20
Percent Similarity: 44.749 Percent Identity: 19.178

alignment_block:
US-08-656-811A-1 x PCT-US95-10202-21 ..

Align seg 1/1 to: PCT-US95-10202-21 from: 1 to: 4249

13 AlaArgGluTrpGlyLeuGluMetProValGlnThrAspGlyGlnPh 29
:::||||| :::::||||| :::::|||||
2048 ACTCGGAAACAACATACATAAGTGAAGAAGTGAACATTGATGCTCAGTA 2097-
29 eGlyAspLeuLysSerThr..... 35
: |||||:|||||
2098 T.....TTGCAGATACATTCAAGGTTTCAAAGCGCCAGTCATTTGCTC 2141
36SorArg 37
2142 CGTTTTCAAATCCAGGAATGCAGAGAGGAGATGCAACATTTCTCTGCC 2191
38 HisGlyGlyAspGluSerLeuGlnProGlnGlyAlaThrLeuLys 54
|||||:|||||: ||| :::::|||||
2192 CACTCTGGG.....TCCTTAAAGAAACAAGTCCAAAGTCACTTTTCA 2235
54 sLeuGluProPheGluGluAspValLeuGlyAlaGluTrpMetGluSerS 71
: ||| |||||:
2236 ATGTGAACAAAAGGAAGAAAT..... 2257
71 erAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeu 87
2258CAAGGAAGAAATGAGTCTAATATC 2281
88 HisProPheGluSerAsnLeuLeuGluPheThrSerLeuIleThrProAs 104
||| :::::
2282 AAGCCTGTACACACA..... 2296
104 pAspSerThrValSerLysAspIleLeuSerSerThrLeuGlnPheProt 121
::: ||| |||||
2297GTTAATATCACTGCAGGCTTTCCCTG 2321
121 hr.....GlnProValAsnIleProLeuTyrAlaSerHis 132
: :::::|||||: |||
2322 TGGTTGGTCAGAAAGATAAGCCAGTTGATAATGCCAAATGTAGTATCAAA 2371
133 GlyAlaGluAspPheSerAlaGluThrGluPhe.....GluAsnH1 146
|||||: ||| :::::|||||
2372 GGAGGCTCTAGGTTTGTCTATCATCTCAGTTCAGAGCGCAACGAACTGG 2421
146 sLeuSerProProAsp.....SerProGluGlnValA 157

2422 ACTCATTTACTCCAAATAAACAATGGACTTTTACAAAACCCATATCGTGATAC 2471
 157 laProValIIEAsnLeuGluProValGluLeuThrAlaSerHisMetThr 173
 2472 CACCACCTTTTCCCATCAAGTCATTTGTTAAACACTAAATGTACAGAAAAT 2521
 174 ValIle.....SerProAspGlyLeuLe 181
 2522 CTGCTGAGAGAAAACCTTTGAGGAACATTTCAATGTGCACCTGAAAGAGAAAT 2571
 181 uGlyGlyMetGluLeuAlaSerGluSerLeuThr..... 192
 2572 GGGAAATCAGAACATTCCAAGTACAGTCAGTCAGCACAAATTAGCGGTAAATAACA 2621
 193PheThrGluLeuAspPheValAsnAsp..... 203
 2622 TTAGAGAAAATGTTTTTAAAGAAGCCAGCTCAAGCAATATTAAATGAAGTA 2671
 204SerAlaValGlySerIleGlyClyAlaGluLeuLeuGlySerPr 219
 2672 GGTTCACGACTACTAATGAAGTGGGTCCAGCTAATTAATGAAATAGGTCC... 2719
 219 OeuSerValAspValGluSerThrIleSerPheSer...GlyProS 235
 2720AGTCATGAAAACATTTCAAGCAGAACTAGGTAGAAACACAGAGGCCAA 2765
 235 er.....SerProGluThrSerGln 241
 2766 AATTGAATGCTATGCTTAGATTAGGGGTTTTGCAACCTGAGGTCTATAAA 2815
 242 SerSerIleIleGluSerSer.....ProGluLeuThrLys..... 253
 2816 CAAAGCTCTTCGGNAGTAATGTAAAGCATCTCGAATAAANAAGCAAGA 2865
 254ValIleSerThrSerSerIleAspAlaSerLysArgPheS 267
 2866 ATATGAAGAAGTAGTTCAGACTGTTAATACAGAT.....TTCT 2903
 267 erProThrSerArgSerSerLysSerLysGlnSerValLysThrSerAsp 283
 2904 CTCCATATCTGATTCAGATAACTTAGAACAGCCTATGGGAAGTAGTGCAT 2953
 284 AlalysAlaProArgLysThrArgThrProAlaGlnProValProGluHi 300
 2954 GCATCT.....CAGGTTTGTTCTCAGACACCTGATGA 2985
 300 svalIleMetGluHisLeuAspLysLysAspArgLysLysLeuGlnAsn. 316
 2986 CTGTTAGATAGTGGTGAATAATAAAGGAAGTACTAGTTTTTCCTGCAAAATG 3035
 317LysAsnAlaAlaIleArgThrArgMetLysLysGlyGlu 330
 3036 ACATTAAGGAAGTCTGCTGTTTTTACGAAAACGCTCCAGAAAGGAGAG 3085
 331AlaGlnGlyIrl 334
 3086 CTTAGCAGGAGTCTCTAGCCCTTTCCACCATACACATTTGCTCGCTGAGGTTA 3135
 334 eLysGlyGluGluGlnLeuGluLeuAsnThrLysLeuLysThrL 351
 3136 CGAAGAGGGGCCAAGAAATTAGAGTCTCTCAGAAGAGAACTTATCTAGTG 3185
 351 ysValAspLeu 355
 3186 AGGATGAAGAGCTT 3199

```
seq_name: /cgn2_6/ptodata/2/1na/PCTUS_COMB.seq:PCT-US95-10203-21
seq_documentation_block:
; Sequence 21, Application PC/TUS9510203
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
```

```

: APPLICANT: Goldgar, David E.
: APPLICANT: Miki, Yoshio
: APPLICANT: Swenson, Jeff
: APPLICANT: Kamb, Alexander
: APPLICANT: Harshman, Keith D.
: APPLICANT: Shattuck-Eidens, Donna M.
: APPLICANT: Tavtiglian, Sean W.
: APPLICANT: Wiseman, Roger W.
: APPLICANT: Futreal, P. Andrew
: TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
: TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
: TITLE OF INVENTION: Susceptibility Gene
: NUMBER OF SEQUENCES: 85
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
: STREET: 1201 New York Avenue, N.W., Suite 1000
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/10203
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US
: FILING DATE: 07-JUN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/409,305
: FILING DATE: 24-MAR-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/348,824
: FILING DATE: 29-NOV-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08-308,104
: FILING DATE: 16-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/300,266
: FILING DATE: 02-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/289,221
: FILING DATE: 12-AUG-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Ihnen, Jeffrey L.
: REGISTRATION NUMBER: 28,957
: REFERENCE/DOCKET NUMBER: 24884-109347
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-962-4810
: TELEFAX: 202-962-8300
: INFORMATION FOR SEQ ID NO: 21:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4249 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
: PCT-US95-10203-21

```

```
alignment_scores:
  Quality: 118.50
  Ratio: 0.605
  Percent Similarity: 44.749
  Length: 438
  Caps: 20
  Percent Identity: 19.178

alignment_block:
```



```

US-08-656-811A-1 x PCT-US95-10203-21
Align seg 1/1 to: PCT-US95-10203-21 from: 1 to: 4249
13 AlaArgGluTrpGlyLeuGluMetProValGlnThrAspGlyGlnPh 29
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2048 AGTCGGGAAACAAGCATAGAAATGGAAGAAAGTGAACACTTGATGCTCAGTA 2097
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
29 eGlyAspLeuLysSerThr.
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2098 T.....TTGCAGATACATTCAAGGTTTCAAGCGCCAGTCATTGGCTC 2141
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
36 .....SerArg 37
2142 CGTTTCAATCCAGGAATGCGAAGAGGAATGTCACACATTCTCTGCC 2191
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
38 HisGlyGlyAspGluSerLeuGlnProGlnGlyAlaThrLeuLys 54
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2192 CACTCTGGG.....TCCCTTAAAGAAACAAGTCCAAAGTCACTTTTGA 2235
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
54 sLeuGluProPheGluGluAspValLeuGlyAlaGluTrpMetGluSerS 71
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2236 ATGTGAACAAAGGAAGAAAT..... 2257
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
71 erAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeu 87
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2258 .....CAAGGAAAGAAATGAGTCTAATATC 2281
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
88 HisProPheGluSerAsnLeuGluPheThrSerLeuIleThrProAs 104
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2282 AAGCCTGTACAGACA..... 2296
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
104 pAspSerThrValSerLysAspIleLeuSerSerThrLeuGlnPheProt 121
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2297 .....GTTAATATCACTGCAGCGCTTCTCCTG 2321
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
121 hr.....GlnProValAsnIleProLeuTyrrAlaSerHis 132
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2322 TGGTTGGTCAGAAAGATAAGCCAGTTGATAATGCCAAATGTAGTATCAAA 2371
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
133 GlyAlaGluAspPheSerAlaGluThrGluPhe.....GluAsnHis 146
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2372 GGAGGCTCTAGTTGTCTATCACTCACTCAGTTCAGAGGCAACGAAACTGG 2421
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
146 sLeuSerProProAsp.....SerProGluGlnVala 157
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2422 ACTCATTTACTCCAAATAAACAATGGACTTTTACAAAACCCATATCGTATAC 2471
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
157 laProValIleAsnLeuGluProValGluLeuThrAlaSerHisMetThr 173
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2472 CACCACTTTTCCCACTCAAGTCATTGTTTAAAACTAAATGTAGAAATAAT 2521
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
174 ValIle.....SerProAspGlyLeuLe 181
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2522 CTGCTAGAGAAACCTTTGAGGAACATTCATGTCACCTGAAAGAGAAAT 2571
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
181 uGlyGlyMetGluLeuAlaSerGluSerLeuThr..... 192
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2572 GGAATATGAGAACATTTCCAAGTACAGTGCAGCACAAATAGCCGTAATAACA 2621
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
193 .....PheThrGluLeuAspPheValAsnPheAsnAsp... 203
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2622 TTAGAGAAATATGTTTTTAAAGAGCCAGCTCAAGCAATATTAATGAAGTA 2671
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
204 ...SerAlaValGlySerIleGlyGlyAlaGluLeuLeuGlySerPr 219
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2672 GGTTCAGTACTAATGAAGTGGCTCCAGTATTATGAATAGGTTC... 2719
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
219 oLeuSerValAspAspValGluSerThrIleSerPheSer...GlyProS 235
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2720 ....ACTGATGAAACATTCACAGCAGCACTAGGTAGAACACAGAGGCCAA 2765
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
235 er.....SerProGluThrSerGln 241
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

```

:::  :::  :::  :::  :::  :::  :::  :::  :::  :::
2766 AATTGAATGCTATGCTTAGATTAGGGGTTTTCACACTGAGGCTCTATAAA 2815
   :::  :::  :::  :::  :::  :::  :::  :::  :::  :::
242 SerSerIleIleGluSerSer.....ProGluLeuTyrrLys.... 253
   :::  :::  :::  :::  :::  :::  :::  :::  :::  :::
2816 CAAAGCTCTCTCTGGAAGTAATTCTAAGCATCCTGAAATAAAAAAGCAAGA 2865
   :::  :::  :::  :::  :::  :::  :::  :::  :::  :::
254 .....ValIleSerThrSerSerIleAspAlaSerLysArgPheS 267
   :::  :::  :::  :::  :::  :::  :::  :::  :::  :::
2866 ATATGAAGAAGTAGTCTCAGACTCTTAATACAGAT.....TTCT 2903
   :::  :::  :::  :::  :::  :::  :::  :::  :::  :::
267 erProTyrrSerArgSerSerLysGlnSerValLysThrSerAsp 283
   :::  :::  :::  :::  :::  :::  :::  :::  :::  :::
2904 CTCATATCTGATTTCAGATACTTAGACAGCCTATGGGAAGTAGTCAT 2953
   :::  :::  :::  :::  :::  :::  :::  :::  :::  :::
284 AlaIleSalalProArgLysThrArgThrProAlaGlnProValProGluH 300
   :::  :::  :::  :::  :::  :::  :::  :::  :::  :::
2954 GCATCT.....CAGGTTTGTCTCAGACACCTGATGA 2985
   :::  :::  :::  :::  :::  :::  :::  :::  :::  :::
300 sValIleMetGluHisLeuAspLysLysAspArgLysLysLeuGlnAsn. 316
   :::  :::  :::  :::  :::  :::  :::  :::  :::  :::
2986 CCTGTTAGATGATGTTGAAATAAAGGAAGATACTAGTTTGTGCTGAAATG 3035
   :::  :::  :::  :::  :::  :::  :::  :::  :::  :::
317 .....LysAsnAlaIleAlaArgTyrrArgMetLysLysGlyGlu 330
   :::  :::  :::  :::  :::  :::  :::  :::  :::  :::
3036 ACATTAAGGAAAGTTCTGCTGTTTGTAGCAAAAGCGTCCAGAAAGGAGAG 3085
   :::  :::  :::  :::  :::  :::  :::  :::  :::  :::
331 .....AlaGlnGlyI 334
   :::  :::  :::  :::  :::  :::  :::  :::  :::  :::
3086 CTTAGCAGGAGCTCTAGCCCTTTCACCCCATACACATTTCCTCAGGGTTA 3135.
   :::  :::  :::  :::  :::  :::  :::  :::  :::  :::
334 eLysGlyGluGluGlnGluLeuGluLeuAsnThrLysLeuLysThrL 351
   :::  :::  :::  :::  :::  :::  :::  :::  :::  :::
3136 CCGAAGAGGGGCCAAGAAATAGAGTCCCTCAGAAGAGAACTTATCTAGTG 3185
   :::  :::  :::  :::  :::  :::  :::  :::  :::  :::
351 ysValAspAspLeu 355
   :::  :::  :::  :::  :::  :::  :::  :::  :::  :::
3186 AGGATGAAGAGCTT 3199
   :::  :::  :::  :::  :::  :::  :::  :::  :::  :::
seq_name: /cgn2_6/ptodata/2/1na/PCTUS_COMB.seq: PCT-US95-10220-21
seq_documentation_block:
; Sequence 21, Application PC/TUS9510220
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: Method for diagnosing a
; TITLE OF INVENTION: Predisposition for Breast and Ovarian Cancer
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10220
; FILING DATE:
; CLASSIFICATION:

```


[illegible]

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2000, 23:32:01 ; Search time 36.41 Seconds
(without alignments)
644.173 Million cell updates/sec

Title: US-08-656-811a-1
Perfect score: 1928
Sequence: 1 MELDLWSEDFQLAREWGLEM.....KYMKNLMEDVCKAKGIQLKM 379

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_64:**

1: pir1:**

2: pir2:**

3: pir3:**

4: pir4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	264	13.7	351	2 A45377	transcription fact
2	248	12.9	381	2 S26812	transcription fact
3	163	8.5	68	2 D34223	transcription fact
4	148.5	7.7	488	2 T27532	hypothetical prote
5	145	7.5	436	2 A37953	transcription regu
6	143.5	7.4	208	2 T24446	hypothetical prote
7	138.5	7.2	389	1 A39429	cAMP response elem
8	137	7.1	358	2 C42026	cyclic AMP respons
9	137	7.1	456	2 B42026	cyclic AMP respons
10	137	7.1	505	1 S05380	cAMP response elem
11	135	7.0	486	2 JC4028	activating transcr
12	134.5	7.0	330	2 T38374	probable bzip tran
13	134	7.0	647	2 S26386	transcription fact
14	133	6.9	448	2 A42026	cAMP response elem
15	133	6.8	483	2 S12741	transcription fact
16	130.5	6.8	582	2 S10099	transcription fact
17	129.5	6.7	313	2 A34785	DNA-binding protei
18	129.5	6.7	457	2 T21063	hypothetical prote
19	127	6.6	181	2 A54023	transcription fact
20	127	6.6	460	2 T19062	hypothetical prote
21	127	6.6	638	2 B35816	transcription regu
22	126.5	6.6	338	2 A53066	CCAAT enhancer-bin
23	126	6.5	699	2 T09069	probable cAMP-resp
24	126	6.5	938	2 A56731	chromatin assemb
25	124.5	6.5	2774	2 A43359	microtubule-associ
26	122.5	6.4	222	2 C34223	transcription fact
27	121	6.3	424	2 T10985	regulator protein
28	121	6.3	721	2 A33319	microtubule-associ
29	121	6.3	728	2 A81385	probable ATP /GTP

30	120.5	6.2	654	2 A34734	transcription fact
31	119	6.2	341	2 S67641	hypothetical prote
32	119	6.2	439	2 JC1178	transforming prote
33	119	6.2	853	2 T46347	hypothetical prote
34	119	6.2	1528	2 JS0703	DNA topoisomerase
35	118.5	6.1	590	2 S66956	hypothetical prote
36	118.5	6.1	1151	2 T24541	hypothetical prote
37	118.5	6.1	1863	1 A58881	breast/ovarian can
38	118	6.1	378	2 T12621	Dc3 promoter-bindi
39	118	6.1	547	2 T40314	hypothetical colle
40	118	6.1	551	2 S67788	hypothetical prote
41	117.5	6.1	427	2 S53021	G-box-binding prot
42	117.5	6.1	682	2 A42141	transcription fact
43	117	6.1	355	2 T40132	ATF/CREB-family tr
44	117	6.1	1230	2 T22458	hypothetical prote
45	117	6.1	1790	2 S67593	transport protein

ALIGNMENTS

RESULT 1
A45377
transcription factor CREB-2 - human
N:Alternate names: TAXREB67
C:Species: Homo sapiens (man)
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 05-Nov-1999
C:Accession: A45377; I56787
R:Karpinski, B.A.; Morle, G.D.; Huggenvik, J.; Uhler, M.D.; Leiden, J.M.
Proc. Natl. Acad. Sci. U.S.A. 89, 4820-4824, 1992
A:Title: Molecular cloning of human CREB-2: an ATF/CREB transcription factor that can
A:Reference number: A45377; MUID:92279218
A:Accession: A45377
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-351 <KAR>
A:Cross-references: GB:M86842; NID:g181040; PIDN:AAA52071.1; PID:g181041
A:Note: the sequence is not complete
R:Tsujiimoto, A.; Niyonoya, H.; Morita, T.; Sato, T.; Shomotohno, K.
J. Virol. 65, 1420-1426, 1991
A:Title: Isolation of cDNA for DNA binding proteins which specifically bind to TAX-re
A:Reference number: I56787; MUID:91140735
A:Accession: I56787
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-351 <RES>
A:Cross-references: GB:D90209; NID:g220087; PIDN:BAA14234.1; PID:g220088
C:Genetics:
A:Gene: GDB:CREB2; ATF2; CRE-BP1; TREB7
A:Cross-references: GDB:I28011; OMIM:123811
A:Map position: 2q32-2q32
C:Keywords: transcription factor

Query Match 13.7%; Score 364; DB 2; Length 351;
Best Local Similarity 28.7%; Pred. No. 1.3e-09;
Matches 117; Conservative 53; Mismatches 122; Indels 116; Gaps 22;

QY	19	EMPVVQTDGQFGDLKS---TSRHGDESLJL-----OPOGAT-----L 53
Db	3	EMSLUSLEVGLDMSPPDPSPGLGAEISGLDDLVLEVAKHFKPHGFSDDKAKASSEWL 62
QY	54	KLEPF-----ERDVLGAEMW-ESSDLGSF-LDALGDNNHRLHPFESNLFTSLIT- 102
Db	63	AVDGLVSPNSNKEDAFSGTDWMLKMDLKEFDLDA-----LLGIDDLFTM 108
QY	103	PDDSTVSKD-----ILSSTLQFPQTPQVINIPLYASHGAEDFSAETEFENHLSPPDS-- 152
Db	109	PDDLTLTLDLDDCDLFAPLVQETNKPQPTVN-PI-----GHL--PESIT 149
QY	153	-PEQVAPVINLEPVELTASHMTVISPDLGLGGMELASLSLFTFELDFVNFNDSAVGSGTG 211
Db	150	KPDQVAPFTFLQPLPLSPGVLS-STPDHSF-SLELGSF-SLELGSF-VDITEGDRKPDYAYVAMI-- 204

[illegible]

A:Molecule type: DNA
A:Residues: 1-488 <EWL>
A:Cross-references: EMBL:Z77136; PIDN:CAR00883.1; GSPDB:GN00023; CESP:ZC376.7
A:Experimental source: clone ZC376
C:Genetics:
A:Gene: CESP:ZC376.7
A:Map position: 5
A:Introns: 32/3; 56/2; 181/3; 205/3; 232/3; 450/2

Query Match 7.7% Score 148.5; DB 2; Length 488;
Best Local Similarity 24.1%; Pred. No. 0.039;
Matches 91; Conservative 53; Mismatches 115; Indels 119; Gaps 21;

QY 21 PVVTDGQ--FGDLKSTRRHGGDESLQPQGATLKLEPFEDV---LGAEW-----M 68
 |: | : ||| : |: | : ||| : |: | :
Db 184 PLFQSPSKSAIDQLTGTSR---IDEYGMPPQDR--KLKSFEMDIEQESKAVDWEAWNHYL 238

QY 69 ESSDLGSFLDALGNHRLHPFNENLLFTSLITPDDSTVSKDILSSTLQPTQPVN--- 125
 |: | : ||| : |: | : ||| : |: | : ||| :
Db 239 ESDD-----DVFKRPEAF---FKEEPIMTSSD-----LMTSSTSPDGSGISLYD 281

QY 126 --IPLYASH-CAEDFSAEETEPN--HLSPDPSP--EQAVPVINLEVPVELTASHTVISP 177
 || :|| : ||| : |: | : ||| : |: | : ||| :
Db 282 PMIPPPSHFPFNLSSSSSNLRLLSTPSAPMQQEHRAVPVRMHHDVDFLSGGLCVCP 341

QY 178 DGLLGGMLASESLTFELDFPNFNDSAVGIGGAELLGSPLSVDVFESTIFSFGSPSP 237
 |: | : ||| : |: | : ||| : |: | : ||| :
Db 342 -----KQEDVFD--DFIQRD-----DDDEDYI----PAS- 365

QY 238 ETSOSIIIESPELYKVISTSIDASKRFSPYSRSSKSQSVKTSDAKAPKRTTPAQP- 296
 |: | : ||| : |: | : ||| : |: | : ||| :
Db 366 -----EARTSSLNRKRSATPTYLRRRDSF---RSWTPTSDD 399

QY 297 -VPEH-----VIMEHLDKK--DRKKLOKNKAIRYRMKKKGAEAGIKGEEOELELN 345
 || :| : ||| : |: | : ||| : |: | : ||| :
Db 400 YFPEHQKFKRGVVLFPSVDVEETDRRLMLNRIAAVRYREKKRAEKKGKKEFEVADRN 459

QY 346 TKLTKVDDLDQREIKYM 363
 | : | : ||| :
Db 460 RILLQKERQLKREINSMK 477

RESULT 5
A37953
transcription regulator PAN-1 - golden hamster (fragment)
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 06-Dec-1991 #sequence_revision 06-Dec-1991 #text_change 02-Jul-1998
C:Accession: A37953
R:German, M.S.; Blanan, M.A.; Nelson, C.; Moss, L.G.; Rutter, W.J.
Mol. Endocrinol. 5, 292-299, 1991
A:title: Two related helix-loop-helix proteins participate in separate cell-specific control of transcription factor 3
A:Reference number: A37953; MUID:91246228
A:Accession: A37953
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-436 <GER>
C:Superfamily: human transcription factor 3

	Query Match	7.5%	Score 145;	DB 2;	Length 436;
	Best Local Similarity	23.1%	Pred. No. 0.055;		
	Matches	86;	Conservative	56;	Mismatches 171; Indels 60; Gaps 14;
QY	27	QGFCDLKSTR-----HGGD-----ESLSLP-----QGATLKLEPFE--EDVLGAEMW	68		
		: : : : : : : :			
Db	46	GTFGLOQQERMSYQLHGSEVNGTLPVSSFSAAPGTGYGGASGHTPPVSGADSLMGTRGT	105		
QY	69	ESSDLGFLDAGNHERLHPFEENLLETSIIIPDDSTVSKDILSLSTLPQPTQPVNIPL	128		
		: : : : : : : : : : : :			
Db	106	TASSSG---DAIGKALASISYFDHSSNNFSP--SFSTPVSGPQGLPGTSQMPRACAPSA	160		
QY	129	YASHGAEFDSAEETEFENHLSPDPSPEQVAPVNIPLNEPVELTASHMTVISPD--GLIGGMEL	186		

```

Db 161 SPTYDGLHGLQSMEDRLD-----EAIHVLRSHAVGTASDLHGLLPG--- 203
QY 187 ASELTTETELDFVNFNDVAVSGSIGGA--BELLGSPLSVDDVESTISFSGPSPETSSQSI 244
Db 204 -HGALTTSFPGPVLPGGRHAGLVGGGHPEDGLTSGTSLTHASLPSQASSLPDJSQ--- 259
QY 245 IESSPELYKVISTSSIDASKRRFSYRSRKK-SKQSVKTSDA-----KAPRKTRTPAOP 296
Db 260 --RPPDSYVGLGRAGAPAGA--SEIKREKKDDEESTSVADAEEDKDLKAPRTRTSPDES 315
QY 297 VPEHVIHEHLDDKDRKKLQNNKNAIRYRMKKKGAGGIGEEQELNKKTKTKVDDLIQ 356
Db 316 TDEVLSLEEKDLRDRERRMANNARERYVRVDINEAFRELGRICQLHLKSDKAQTLLIIQ 375
QY 357 REIKYMKNLMDV 369
Db 376 QAVQVILGLEQV 388

RESULT 6
T24446
hypothetical protein T04C10.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T24446
R:Burton, J.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z19891
A:Accession: T24446
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-208 <WTL>
A:Cross-references: EMBL:Z69885; PIDN:CAA93757.1; GSPDB:GN00028; CESP:T04C10
A:Experimental source: clone T04C10
C:Genetics:
A:Gene: CESP:T04C10.4
A:Map position: X
A:Introns: 91/3; 183/3

```

Query Match	7.4%;	Score 143.5;	DB 2:	Length 208;
Best Local Similarity	27.0%;	Pred. No. 0.024;		
Matches 43;	Conservative 28;	Mismatches 69;	Indels 19;	Gaps
Qy	230	SFSGSPSPETSSQSIIESSPELYKVI	TSST--DASKRSPYRSRSKSKSQSVK	TSDAK-- 285
		: : : : : : : :		
Db	43	SYFNPYHQSYQQHLLNSDYNFQW	POSTSPVPPDFCSIEPMTNVOAQE	ILEIVREC 102
		: : : : : : : :		
Qy	286	--APRKTRTPAQVPPEHVI	MEH-----LDDKKRKLQKNKNAAI	IYRMMKKGE 330
		: : : : : : : :		
Db	103	EEIERRSNSSASPASNWSDEHSD	OSEKSYHPYKTPPEKKERRKAQNL	LAATRYREKKRE 162
		: : : : : : : :		
Qy	331	AQGTGGEQELEEELNTKLTKY	VDLQREIKYMKNLMDV	369
		: : : : : : : :		
Db	163	KEEAMTCIEGLSVNGLKDOQV	SELEIREIYFYKFKFTEM	201
		: : : : : : : :		

RESULT 7
A39429
cAMP response element-binding protein ATF2 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A39429
R:Kageyama, R.; Sasai, Y.; Nakanishi, S.
J. Biol. Chem. 266, 15525-15531, 1991
A:Title: Molecular characterization of transcription factors that bind to the
A:Reference number: A39429; MUID:91332085
A:Accession: A39429
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-389 <RAG>
A:Cross-references: GB:M65148; NID:g2065569; PID:AAAA2013.1; PID:g2065570

C: Superfamily: cAMP response element-binding protein 1; fos/jun DNA-binding domain homolog
C: Keywords: DNA binding; nucleus; transcription regulation
F: 231-271/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 7.2%; Score 138.5; DB 1; Length 389;
Best Local Similarity 23.6%; Pred. No. 0.12;
Matches 77; Conservative 47; Mismatches 137; Indels 65; Gaps 14;

QY 82 DNERHLPFESNLLEFSLTP--DDSTVSKDILSLTQPTQTPVNIPLY---ASHGAE 136
DB 21 EDHLAVHKHHK---EMTLKFGPARNDVIVADQPTTRFLKNCSEVGLFNELASPFENE 77
QY 137 FSAETEFENHLSPPDSEQVAPVINL---EPVELTASHMTVISPDLGGMELASESLTF 193
DB 78 FKRASEDDIKMPLDLSPLATPIIRSKIEEPSVVEITHQDSPLPH-----PESTT- 127
QY 194 TELDFVNFDSAVGSGAELGLSPUSVDVESTISFSGPSPSPETSSQSSIISSPELYK 253
DB 128 -----ND-----EKLVRPVTM--VPSVPGIPGSPSPQVQS---EAKMRLKA 164
QY 254 VIS-----TSSIDASKRFSYSSKSKQSVKTSDAKAPKRTTPAQP---VPE----- 299
DB 165 ALTQQHPVTVNGDTVKGHSGLVRAOSEERPOSLOQPATSTTETPASPAHTTPQTQNTS 224
QY 300 -----HVIMEHLKDKRKKLQ--NKNAAIRYRMKKKGAEQIGKEQELEELNTKTKVDD 354
DB 225 GRRRAANEDPDEKRRKFLERNRAASRCRQKRVWVQSLKKAEDLSLNGQLQSEVTL 284
QY 355 LQREIKYMKMLM--EDVCKAKGIQLK 378
DB 285 LRNEVAQLKQLLAHKDCPVTAMQK 310

RESULT 8
C42026
cyclic AMP response element DNA-binding protein isoform 2 - mouse
C: Species: Mus musculus (house mouse)
C: Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 26-Aug-1999
C: Accession: C42026

R: Georgopoulos, K.; Morgan, B.A.; Moore, D.D.
Mol. Cell. Biol. 12, 747-757, 1992
A: Title: Functionally distinct isoforms of the CRE-BP DNA-binding protein mediate activation
A: Reference number: A42026; MUID: 92123199

A: Contents: EL4
A: Accession: C42026
A: Status: preliminary
A: Molecule type: mRNA
A: Residues: 1-358 <GEO>
A: Cross-references: GB:S76659; NID:G243430; PIDN:AAB21129.1; PID:G243431
A: Note: sequence extracted from NCBI backbone (NCBIN:76659, NCBI:P:76660)
C: Superfamily: cAMP response element-binding protein 1; fos/jun DNA-binding domain homolog
F: 199-239/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 7.1%; Score 137; DB 2; Length 358;
Best Local Similarity 23.2%; Pred. No. 0.13;
Matches 70; Conservative 45; Mismatches 127; Indels 60; Gaps 11;

QY 104 DSTVSKDILSLTQPTQVNIPLY---ASHGAEFSAETEFENHLSPPDSEQVAPVI 160
DB 10 NDSVIVADQPTTRFLKNCSEVGLFNELASPFENEKASEDDIKMPLDLSPLATPII 69
QY 161 NL---EPVELTASHMTVISPDLGGMELASESLTFELDFVNFDSAVSIGGAELLG 217
DB 70 RSKIEPSVVVETTHQDSPLPH-----PESTT-----SDKLV 101
QY 218 SPLSVDDVESTISFSGPSPSPETSSQSSIESPELYKVIS-----TSSIDASKRFSYSR 271
DB 102 RPYTM--VPSVPGIPGSPSPQVQS---EAKMRLKALTOQHPPVTVNGDIVKGHSGSLVR 156
QY 272 SSKSKQSVKTSDAKAPKRTTPAQP---VPE-----HVIMEHLKDKRKKLQ--NKN 318

DB 157 TQSEESRQSLQOPATSTTETPASPAHTTPQTQNTSGRRRAANEDPDEKRRKFLERNRA 216
QY 319 AAIRYRMKKKGAEQIGKEQELEELNTKTKVDDIQRKIKYMKMLM--EDVCKAKGIQ 376
DB 217 AASRCRQKRVWVQSLKKAEDLSLNGQLQSEVTLRNEVAQLKQLLAHKDCPVTAMQ 276
QY 377 LK 378
DB 277 KK 278
RESULT 9
B42026
cyclic AMP response element DNA-binding protein isoform 1 - mouse
C: Species: Mus musculus (house mouse)
C: Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 26-Aug-1999
C: Accession: B42026
R: Georgopoulos, K.; Morgan, B.A.; Moore, D.D.
Mol. Cell. Biol. 12, 747-757, 1992
A: Title: Functionally distinct isoforms of the CRE-BP DNA-binding protein mediate act
A: Reference number: A42026; MUID: 92123199
A: Contents: EL4
A: Accession: B42026
A: Status: preliminary
A: Molecule type: mRNA
A: Residues: 1-456 <GEO>
A: Cross-references: GB:S76657; NID:G243428; PIDN:AAB21128.1; PID:G243429
A: Note: sequence extracted from NCBI backbone (NCBIN:76657, NCBI:P:76658)
C: Superfamily: cAMP response element-binding protein 1; fos/jun DNA-binding domain ho
F: 297-337/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 7.1%; Score 137; DB 2; Length 456;

Best Local Similarity 21.7%; Pred. No. 0.19;
Matches 85; Conservative 57; Mismatches 168; Indels 82; Gaps 15;

QY 52 TKLLEPFEED-VLGAEMWESSDLGSLF---DALGDNHERLHPFESNLLEFTSLITPDOST 107
DB 2 TLKFGPARNDVIVAD--QTPTTRFLKNCSEVGLFNELASPFEN---EFKASEDDIKK 56
QY 108 VSKDILSLTQPTQVNIPLYASHGAEDFSAETEFEN----HLSPDPSPEQVAPVIN-- 161
DB 57 MPLDL-----SPLATPIIRSKIEEPSVVEITHQDSPLPHPESTTSDEKEYPLAQTA 107
QY 162 -----LEPVELTASHMTVISPDLGGMELASESLTFELDFVNFDSAVSIGGAEL 215
DB 108 QPSTAIRPASLOQPNVLLTSSDSSVLIQQAIVPSTSTVITQAPSSNRPIVPVPGPFPL 167
QY 216 L-----GSPLSVDD-----VESTISFSGPSPSPETSSQSSIIES 247
DB 168 LLHLPLNGQTPVAIPASITSSNVHVPAAVPLVRPVTWVPSVPGIPGSPSPQVQS---EA 224
QY 248 SPELYKVIS-----TSSIDASKRFSYSSKSKQSVKTSDAKAPKRTTPAQP---VP 298
DB 225 KMRLKALTOQHPPVTVNGDTVKGHSGLVRTQSESRPSQLQOPATSTTETPASPAHTTP 284
QY 299 E-----HVIMEHLKDKRKKLQ--NKNAAIRYRMKKKGAEQIGKEQELEELNTKL 348
DB 285 QTQNTSGRRRAANEDPDEKRRKFLERNRAASRCRQKRVWVQSLKKAEDLSLNGQL 344
QY 349 KTKVDDIQRKIKYMKMLM--EDVCKAKGIQLK 378
DB 345 QSEVTLRNEVAQLKQLLAHKDCPVTAMQK 376

RESULT 10
S05380
cAMP response element-binding protein 1 - human
N: Alternate names: transcription factor, TGACGCA-binding
N: Contents: cAMP response element-binding protein HB16
C: Species: Homo sapiens (man)
C: Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 29-Sep-1999
C: Accession: S05380; A34776; B34223

R;Maekawa, T.; Sakura, H.; Kanei-Ishii, C.; Sudo, T.; Yoshimura, T.; Fujisawa, J.I.; Ycse
EMBO J. 8, 2023-2028, 1989
A:Title: Leucine zipper structure of the protein CRE-BP1 binding to the cyclic AMP response
A:Reference number: S05380; MUID:90005408
A:Accession: S05380
A:Molecule type: mRNA
A:Residues: 1-505 <MAE>
A:Cross-references: EMBL:X15875; NID:g30214; PIDN:CAA33886.1; PID:g30215
R;Kara, C.J.; Liou, H.C.; Ivashkiv, L.B.; Glimcher, L.H.
Mol. Cell. Biol. 10, 1347-1357, 1990
A:Title: A cDNA for a human cyclic AMP response element-binding protein which is distinct
A:Reference number: A34776; MUID:90205810
A:Accession: A34776
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 211-222, 'N', 224-505 <KAR>
A:Cross-references: GB:M31630; NID:gl83787; PIDN:AAA35951.1; PID:g386762
R;Hai, T.; Liu, F.; Coukos, W.J.; Green, M.R.
Genes Dev. 3, 2083-2090, 1989
A:Title: Transcription factor ATF cDNA clones: an extensive family of leucine zipper proteins
A:Reference number: A91622; MUID:90185187
A:Accession: B34223
A>Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 107-357, 'V', 359-465 <HA2>
C:Genetics:
A:Gene: GDB:CREB2; ATF2; TREB7; CRE-BP1
A:Cross-references: GDB:128011; OMIM:123811
A:Map position: 2q32-2q32
C:Superfamily: cAMP response element-binding protein 1; fos/jun DNA-binding domain homolog
C:Keywords: alternative splicing; DNA binding; nucleus; transcription regulation
F;347-387/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 7.1%; Score 137; DB 1; Length 505;
Best Local Similarity 21.7%; Pred. No. 0.22;
Matches 85; Conservative 57; Mismatches 168; Indels

QY	52	TLKLEPFEED-VLGAEWMESSDGLSPL----	DALGONHERLHPFENLLEFTSLITPDOST	107
		: : : : : :		
Db	52	TLFAGPARDNSVIVAD--QTPTPTRELKCEEVGLFNELASPFE--	EFFKASEDDIKK	106
QY	108	VSKDELSSLTQFTQTPVNPIYLASHCAEDFSAETEFEN----	HLSPDPSPEQVAPVIN--	161
		: : : : : :		
Db	107	MLPDL-----SPLATPIIRKSTEESPVVETTHQDPSPLPHEPSTTSDEKEVPLAQT	157	
QY	162	-----LEPVELTSHTMTWISPDGLGGMELASESLTFELDFVFNDSAVGSGIGAEEL	215	
		: : : : : : : :		
Db	158	QPTSAIVRASLOQVPNVLTTSSDSSVILIOQAVPSTSSIVITQAPSSNRPIVPVGPPL	217	
QY	216	L-----GSPUSVD-----	VESTISFGSPSPETSQSIIIES	247
		: : : : : :		
Db	218	LLHLPSSGQTMPVAIPASITSSNVHVPAAPVPLVRPVTMPSPGIPGSPSPQVQS--	EA	274
QY	248	SPELYKVIS-----TSSIDAKRRSPYSRSKSKQSVKTS	DAKAPKTRTTPAQP----	298
		: : : : : :		
Db	275	KMLKAAALTOQHPPVTVNGDVTYKGHSGSLVRTQSEESRPOSLOQAPTSTTETPASEAHTPT	334	
QY	299	E-----HVIMEHLDKDKRRKLIQ--KNNAAI	RIYRMKKKGAQGIKEQEOLNKL	348
		: : : : : :		
Db	335	QTOSTSGRRRAANEDPDEKRRKFLERNRAASRCQKRVVQVSLKKAEDLSINGQL	394	
QY	349	KTQVDLQREIKYMKNL-----EDVCKAKGIQLK	378	
		: : : : : :		
Db	395	QSEVITLIRNEVAOLKOLLAHKDCPYTAMOKK	426	

RESULT 11

activating transcription factor 2 - African clawed frog
 N;Alternate names: cyclic AMP-response element-binding protein
 C;Species: Xenopus laevis (African clawed frog)
 C;Date: 13-Jun-1995 #sequence revision 14-Jul-1995 #text change

C/Accession: JC4028
 F/Villarreal, X.C.; Richter, J.D.
 Gene 153, 225-229, 1995
 A/Title: Analysis of ATR2 gene expression during early Xenopus
 A/Reference number: JC4028; MUID:95180723
 A/Accession: JC4028
 A/Molecule type: mRNA
 A/Residues: 1-486 <YIL>
 A/Cross-references: GB:U16158; NID:g887779; PIDN:AAA9518.1
 C/Comment: This protein is a sequence-specific DNA-binding
 C/Superfamily: cAMP response element-binding protein 1; fos;
 C/Keywords: leucine zipper; phosphoprotein; transcription r
 F/37-398/Region: leucine zipper motif
 F/329-369/Domain: fos/jun DNA-binding domain homology <F>
 F/332-364/Region: basic
 F/82/Binding site: phosphate (ser) (covalent) (by protein k
 F/102/Binding site: phosphate (Ser) (covalent) (by protein k

Query Match 7.0%; Score 135; DB 2; Length 486;
Best Local Similarity 21.5%; Pred. No. 0.28;

Qy	52	TLKLEPPEED-VLGAEMWESSDLSGSEL---DALGDNRHLHPESNLLBFTSLITPDDST	107
Db	34	TLKFGPARNDSVIAD--QTPTTRFLKNCEEVGLFENELASPEEN---EFKK-ASEEDKK	87
Qy	108	VSKDILSSTLOFTPQNIPLYASHAGAEDFSATETENHILSPDPDSPEQV-----APVI	160
Db	88	MPDLPL-----SPLATPIIRKSKEEPSVVETTHQD--SPLPHPESTTSQHNQVPCLA	136
Qy	161	N-----LEPVELTSASHMTVTSPOCLUGGMELASLESFTBELDFVNENDSAVGSIGA	212
Db	137	QTAQPSAIYRPASLPQPNVLNTSSDSSVIIQQAVPSPTSSTVITQAPTNSIVTLPQP	196
Qy	213	EELL-----GSLPSVD-----VESTISFGSPSPETSQSISI	244
Db	197	FVILLHLHLPSCGTMPVAIPASITSSNVHVPAAPVLLRPLTMVPSVPCIGPSSPPQVS--	254
Qy	245	IESPPELYKYISTSSIDASKRFSS-----PYSRSKSKSQSVKT	281
Db	255	-----EAKMRFNAAMTQQHPVTNGDTVKHGSGFAITQSEESRPSQLQ	298
Qy	282	SDAKAPKRTR-TPAQVPYE-----HYIMEHLDKDKRKLIQ-NKNAAIRYRMKKKGE	330
Db	299	QPATSTTETRPSAQTPPQTSQSGRRRAANEDPDDEKRSKIQRNRAAASCRCQRKKYW	358
Qy	331	AQGIKGEEQELEEUNTKLKTKYVDDLOREIKYMKNLM--EDVCKAKGIOLK	378
Db	359	VQSLKKKADELISUNGYLQNEVTLTLRNVEAQLKOLLHAHKDCVPTAMOKK408	

RESULT 12
T38374

probable bZip transcription regulator - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 07-Dec-1999
C:Accession: T38374
R:McLean, J.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.; Wood, V.
submitted to the EMBL Data Library, September 1997

A;Accession: J38374
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 1-330 <MCL>
A;Cross-references: EMBL:Z70691; PID:CAA94632.1; G
A;Experimental source: clone c25G10
C;Genetics:
A;Gene: SPDB:SPAC25G10.03
A;Map position: 1

Query Match
7.0%; Score 134.5; DB 2; Length 330;

Search completed: August 8, 2000, 02:39:56
Job time: 11275 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2000, 02:33:46 ; Search time 35.03 Seconds
(without alignments)
335.280 Million cell updates/sec

Title: US-08-656-811A-1
Perfect score: 1928
Sequence: 1 MELDLWSEDFQLAREWGLEM.....KYMKNLMEDVCKAGIQLMK 379

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues

Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_38:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	264	13.7	351	1 ATF4_HUMAN	P18848 homo sapien
2	257	13.3	349	1 ATF4_MOUSE	Q06507 mus musculus
3	142.5	7.4	374	1 FOS_TETFL	Q91496 tetradodon f
4	138.5	7.2	376	1 FOS_FUGRU	P53450 fugu rubrip
5	137	7.1	487	1 ATF2_CHICK	O93602 gallus gall
6	137	7.1	487	1 ATF2_HUMAN	P15336 homo sapien
7	137	7.1	487	1 ATF2_MOUSE	P16951 mus musculus
8	134.5	7.0	330	1 YDC3_SCHPO	Q10424 schizosacch
9	134	7.0	487	1 ATF2_RAT	Q00969 rattus norv
10	134	7.0	647	1 SKO1_YEAST	Q02100 saccharomyc
11	133	6.9	483	1 ATF4_HUMAN	P17544 homo sapien
12	127	6.6	181	1 ATF3_HUMAN	P18847 homo sapien
13	127	6.6	638	1 PAN1_RAT	P21676 rattus norv
14	124.5	6.5	2774	1 MAPA_RAT	P34926 rattus norv
15	121	6.3	721	1 ENP1_TORCA	P14400 torpedo cal
16	120.5	6.2	331	1 AP1_PIG	P56432 sus scrofa
17	120.5	6.2	654	1 TFE2_HUMAN	P34521 caenorhabdi
18	119.5	6.2	893	1 MYC2_CAEEL	P34531 caenorhabdi
19	119	6.2	439	1 MYC_HYLLA	P49033 hylobates l
20	119	6.2	1528	1 TP2A_MOUSE	Q01320 mus musculus
21	118.5	6.1	181	1 ATF3_MOUSE	Q06765 mus musculus
22	118.5	6.1	1863	1 BRCL_HUMAN	P38398 homo sapien
23	118	6.1	551	1 YDC2_YEAST	Q07657 saccharomyc
24	117.5	6.1	682	1 ATF4_HUMAN	O99081 homo sapien
25	117	6.1	1790	1 USO1_YEAST	P25386 saccharomyc
26	116.5	6.0	181	1 ATF3_RAT	P29596 rattus norv
27	116	6.0	334	1 AP1_RAT	P17325 rattus norv
28	116	6.0	3358	1 PGCV_MOUSE	Q62059 mus musculus
29	115.5	6.0	587	1 YID6_YEAST	P40535 saccharomyc
30	115	6.0	334	1 AP1_MOUSE	P05627 mus musculus
31	113.5	5.9	2025	1 TTC3_HUMAN	P53804 homo sapien
32	113.5	5.9	2805	1 MAPA_HUMAN	P78559 homo sapien
33	113	5.9	281	1 TFE1_MOUSE	P15806 mus musculus

34 112.5 5.8 1065 1 SED4_YEAST
35 112 5.8 439 1 MYC_PANTR
36 112 5.8 507 1 YG46_YEAST
37 111.5 5.8 1616 1 P200_MYCGE
38 111.5 5.8 3685 1 DMD_HUMAN
39 111 5.8 211 1 CES2_CABEL
40 111 5.8 2464 1 MAPB_MOUSE
41 110.5 5.7 326 1 FRA2_MOUSE
42 110.5 5.7 331 1 API_HUMAN
43 110.5 5.7 1938 1 MYSA_MOUSE
44 110.5 5.7 1938 1 MYSA_RAT
45 110 5.7 482 1 YSR2_CABEL

ALIGNMENTS

RESULT 1
ATF4_HUMAN

ID ATF4_HUMAN STANDARD; PRT; 351 AA.
AC P18848;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE CYCLIC-AMP-DEPENDENT TRANSCRIPTION FACTOR ATF-4 (DNA-BINDING PROTEIN
DE TAXREB67) (CYCLIC AMP RESPONSE ELEMENT-BINDING PROTEIN 2) (CREB2).
GN ATF4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91140735.
RA Tsujimoto A., Niyunoya H., Morita T., Sato T., Shimotohno K.;
RT "Isolation of cDNAs for DNA-binding proteins which specifically bind
RT to a tax-responsive enhancer element in the long terminal repeat of
RT human T-cell leukemia virus type I.";
RL J. Virol. 65:1420-1426(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92279218.
RA Karpinski B.A., Morle G.D., Huggenvik J., Uhler M.D., Leiden J.M.;
RT "Molecular cloning of human CREB-2: an ATF/CREB transcription factor
RT that can negatively regulate transcription from the cAMP response
RT element.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:4820-4824(1992).
RN [3]
RP SEQUENCE OF 274-341 FROM N.A.
RX MEDLINE; 90185187.
RA Hal T., Liu F., Coukos W.J., Green M.R.;
RT "Transcription factor ATF cDNA clones: an extensive family of leucine
RT zipper proteins able to selectively form DNA-binding heterodimers.";
RL Genes Dev. 3:2083-2090(1989).
RN [4]
RP ERRATUM.
RA Hal T., Liu F., Coukos W.J., Green M.R.;
RL Genes Dev. 4:682-682(1990).
CC -!- FUNCTION: THIS PROTEIN BINDS TO A TAX-RESPONSIVE ENHANCER
CC ELEMENT IN THE LONG TERMINAL REPEAT OF HUMAN T-CELL LEUKEMIA
CC VIRUS TYPE I.
CC -!- FUNCTION: THIS PROTEIN BINDS THE CAMP RESPONSE ELEMENT (CRE)
CC (CONSENSUS: 5'GTGAGCT(A/G)(A/G)-3'), A SEQUENCE PRESENT IN MANY
CC VIRAL AND CELLULAR PROMOTERS.
CC -!- SUBUNIT: BINDS DNA AS A DIMER.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: TO OTHER BZIP PROTEINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation.
The European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

```

CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; D90209; BAA14234.1; -.
DR EMBL; M86842; AAA52071.1; -.
DR PIR; D34223; D34223.
DR HSP; P03069; IDGC.
DR TRANSFAC; T01303; -.
DR MIM; 604064; -.
DR PFAM; PF00170; bZIP; 1.
DR PROSITE; PS00036; bZIP_BASIC; 1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein.
FT DNA_BIND 280 300
FT DOMAIN 306 334
FT CONFLICT 284 284 K -> R (IN REF. 2).
FT CONFLICT 290 290 T -> R (IN REF. 3).
FT CONFLICT 329 331 KEI -> REK (IN REF. 3).
FT CONFLICT 338 338 I -> L (IN REF. 3).
SQ SEQUENCE 351 AA; 38558 MW; 7A708C5CDD6ED7F8 CRC64;

Query Match 13.7%; Score 264; DB 1; Length 351;
Best Local Similarity 28.7%; Pred. No. 3.3e-09;
Matches 117; Conservative 53; Mismatches 122; Indels 116; Gaps 22;

CC 19 EMPVQTDGQFGDLKS---TSRHGGDESLSL-----QPGAT-----L 53
CC ||| : : ||| | | : ||| | : | : | : |
DB 3 EMSFLSEVLVGLDMSFPDPSGLGAESLGLDDYLEVAHKFKPHGFSDDKAKGSSEWL 62
CC ||| : : ||| | | : ||| | : | : | : |
QY 54 KLEFF-----EEDVLGAEMW-EESDLGSF-LDAIGDNHERLHPESNLEFSLIT- 102
CC : : | : | : | : | : | : | : | : | : |
DB 63 AVDGLVSPNSNKEDAFSGTDMLEKMDLKEFDLDA-----LLGIDDLLETM 108
CC : : | : | : | : | : | : | : | : | : |
QY 103 PDDTSVSKD-----ILSTLFQTPQVNIPLYASHGAEDFAETEFENHLSPPDS-- 152
CC ||| : : ||| | | : ||| | : | : | : |
DB 109 PDDLTLTDDTCDFAPLVQETNKQPQTVM-PI-----GHL--PESLT 149
CC : : | : | : | : | : | : | : | : | : |
QY 153 -PEQVAPVINLEPVELFASHMTVISPDLGLGMBELASESLTFTELVNFNDSAVGSI 211
CC ||| : : ||| | | : ||| | : | : | : |
DB 150 KPDQVAPFTFLQPLSPGVLS-STPDHSF-SLELGSE-VDITEGDRKPDYATVAMI-- 204
CC : : | : | : | : | : | : | : | : | : |
QY 212 AEELGSLVDDVESITSGSPSPETSQSSITSESPELYKVVISTSIDASKRFSYSR 271
CC ||| : : ||| | | : ||| | : | : | : |
DB 205 -----PQCIKEED-----TPSDNSGIC-MSPESYL-----GSPQHSSTR 239
CC : : | : | : | : | : | : | : | : | : |
QY 272 SSKSKQSVKYSDAKAPKTRTPAQVPVEHVTM-----EHLDDKDKKLQNKNAAIRYRMK 326
CC ||| : : ||| | | : ||| | : | : | : |
DB 240 GSPNR-SLSPGVLCGSRAPKPYDPGCKVAAKVGKDLKLLKMEQNKTAATRIARQK 298
CC : : | : | : | : | : | : | : | : | : |
QY 327 KKGAGQIKGEEQEELENTKLTKYVDDLQREIKYMKMLMEDYCKAKG 374
CC ||| : : ||| | | : ||| | : | : | : |
DB 299 KRAQEALTGCKEKEKNEALKERADSLAKEIQYLKDLIEVRKARG 346
CC : : | : | : | : | : | : | : | : | : |

RESULT 2
ATF4_MOUSE STANDARD; PRT; 349 AA.
ID ATF4_MOUSE
AC Q06507; Q61906;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYCLIC-AMP-DEPENDENT TRANSCRIPTION FACTOR ATF-4 (C/EBP-RELATED ATF)
DE (C/ATF) (TAXREB67 HOMOLOG).
GN ATF4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-ADIPOCYTE;
RX MEDLINE; 93281642.
RA Valjejo M., Ron D., Miller C.P., Habener J.F.;
RT "c/ATF, a member of the activating transcription factor family of
DNA-binding proteins, dimerizes with CAAT/enhancer-binding proteins

```

```

RT and directs their binding to cAMP response elements.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:4679-4683(1993).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-CD-1; TISSUE-EMBRYO;
RX MEDLINE; 92335183.
RA Chevray P.M., Nathans D.;
RT "Protein interaction cloning in yeast: identification of mammalian
proteins that react with the leucine zipper of Jun.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:5789-5793(1992).
CC -|- FUNCTION: BINDS TO ASYMMETRIC CAMP RESPONSE ELEMENTS (CRE) AS
A HETERODIMER AND TO PALINDROMIC CRE'S AS A HOMODIMER.
CC -|- SUBUNIT: BINDS DNA AS A HOMO- AND HETERODIMER.
CC -|- SUBCELLULAR LOCATION: NUCLEAR.
CC -|- TISSUE SPECIFICITY: BRAIN, HEART, LIVER, SPLEEN, LUNG, AND MUSCLE.
CC -|- SIMILARITY: TO OTHER bZIP PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@lsb-sib.ch).
CC -----
CC EMBL; L13791; AAA40476.1; -.
CC EMBL; M94087; AAA53043.1; ALT_INIT.
CC HSP; P03069; IDGC.
CC MGD; MGI:88096; ATF4.
CC PFAM; PF00170; bZIP; 1.
CC PROSITE; PS00036; bZIP_BASIC; 1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein.
FT DNA_BIND 278 298
FT DOMAIN 304 332
FT CONFLICT 345 345 LEUCINE-ZIPPER (PROBABLE).
FT CONFLICT 345 345 Q -> K (IN REF. 2).
SQ SEQUENCE 349 AA; 38355 MW; 0C3F89574051C7B9 CRC64;

Query Match 13.3%; Score 257; DB 1; Length 349;
Best Local Similarity 27.4%; Pred. No. 8.6e-09;
Matches 115; Conservative 40; Mismatches 123; Indels 142; Gaps 18;

QY 19 EMPVQTDGQFGDLKS---TSRHGGDESLSLQPO---GATLKLEPFEEDYLG-AEW--- 67
DB ||| : : ||| | | : ||| | : | : | : |
QY 68 ----MESSDLGSFLDALGDNHERLHPESNLEFSLITPDDSTVSKDILSSTL----- 117
DB ||| : : ||| | | : ||| | : | : | : |
DB 63 DDGLASASDTGK-EDAFSGTDMLEKMDLKEFDLAFRMDDLTMDPELTLTDDTCDL 121
QY 118 -----OPFTOPVNIPLYASHGAEDFAETEFENHLSPPDS---PEQVAPVINLEP- 164
DB ||| : : ||| | | : ||| | : | : | : |
DB 122 FAPLVQETNKPEPQTVM-PI-----GHL--PESLIKVDQVAPFTFLQPF 162
QY 165 -----VEL-----TASHTVISPDLGLGMBELASESLTFT 194
DB ||| : : ||| | | : ||| | : | : | : |
DB 163 PCSPGVLSSTPEHSFSLSELGSEVDISEGDRKPDAAVITLIP-----CVK 208
QY 195 ELDFVNFNDSAVSIGAEELGSLVSDVESTISFGSPSPETSQSSITSESPELYKV 254
DB ||| : : ||| | | : ||| | : | : | : |
DB 209 EEDTPSDNS--GTCMSPESYLGSP-----QHSPTSR-----APPD--NL 245
QY 255 ISTSSIDASKRFSYRSKSKQSVKTSDAKAPKTRTPAQVPVEHVTMELDKDKRKL 314
DB ||| : : ||| | | : ||| | : | : | : |
DB 246 PSPGSGSGSPRPDY-----PPGVSLTAKVTKELDKRKLKME 284
QY 315 QNKNAAIRYRMKKGAEQKGEQEELENTKLTKYVDDLQREIKYMKMLMEDYCKAKG 374
DB ||| : : ||| | | : ||| | : | : | : |
DB 285 QNKTAATRIARQKRAEQEALTGCKEKEKNEALKERADSLAKEIQYLKDLIEVRKARG 344

RESULT 3
FOS_TETFL

```

```

ID FOS_TETFL STANDARD; PRT; 374 AA.
AC 091496;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE P55-C-FOS PROTO-ONCOGENE PROTEIN (CELLULAR ONCOGENE C-FOS).
GN FOS
OS Tetraodon fluviatilis (Puffer fish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Acanthopterygii;
OC Percomorpha; Tetraodontiformes; Tetraodontidae;
OC Tetraodon.
RN [1]
RP SEQUENCE FROM N.A.
RA Chang M.S., Chang G.D., Huang F.L., Huang C.J., Lo T.B.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NUCLEAR PHOSPHOPROTEIN WHICH FORMS A TIGHT BUT NON-
CC COVALENTLY LINKED COMPLEX WITH THE C-JUN/AP-1 TRANSCRIPTION
CC FACTOR. C-FOS HAS A CRITICAL FUNCTION IN REGULATING THE
CC DEVELOPMENT OF CELLS DESTINED TO FORM AND MAINTAIN THE SKELETON.
CC IT IS THOUGHT TO HAVE AN IMPORTANT ROLE IN SIGNAL TRANSDUCTION.
CC CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).
CC -!- SUBUNIT: HETERODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY. FOS SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U53520; AAB07359.1; -
CC HSSP; P01100; 1FOS.
CC DR PFAM; PF00170; BZIP; 1
CC DR PRINTS; PR00042; LEUZIPPROS.
CC DR PROSITE; PS00036; BZIP_BASIC; 1.
CC KW Proto-oncogene; Nuclear protein; Phosphorylation; DNA-binding.
CC FT DNA_BIND 122 143 BASIC MOTIF.
CC FT DOMAIN 148 176 LEUCINE-ZIPPER.
CC SQ SEQUENCE 374 AA; 40525 MW; 4DE2CB426D9FEEB9 CRC64;

Query Match 7.4%; Score 142.5; DB 1; Length 374;
Best Local Similarity 26.6%; Pred. No. 0.065;
Matches 45; Conservative 35; Mismatches 76; Indels 13; Gaps 4;

Qy 208 SIGGAELLGSLVDDVESTISFGSPSPETSOSIIESSPELYKVIS--TSSIDASKR 265
Db | : : : : : | : : : : : | : : : : : | : : : : : | : :
28 SPAGSYSSMGSPQSQDLTDLTAS----SASFVPTVTAISTSPDLQWVPLVSSVAPSR 83
Qy 266 FSPYSRSSKSKOSVKTSDAKAPKTRTPAQVPVPHVIMEHLDDKDRKLQNKNAIRYRM 325
Db | : : : : : | : : : : : | : : : : : | : : : : : | : :
84 AHPTSPSPYKRTVMRSASPHAKRGV----EQTPEEEKKRIRERNKQAAKCRN 139
Qy 326 KKGGAQGIKEOELEINTKLTKVDDLORE----IKYMKNLMDVCK 371
Db | : : : : : | : : : : : | : : : : : | : : : : : | : :
140 RRRLTDSLQAEITDLEAKSSQLNDIANLKKERLEFILAHHQPICK 188

RESULT 4
FOS_FUGRU
ID FOS_FUGRU STANDARD; PRT; 376 AA.
AC P53450;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE P55-C-FOS PROTO-ONCOGENE PROTEIN (CELLULAR ONCOGENE C-FOS).
GN FOS
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Acanthopterygii;
OC Percomorpha; Tetraodontiformes; Tetraodontidae;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96202283.
RA Trower M.K., Orton S.M., Purvis I.J., Sanseau P., Riley J.,
RA Christodoulou C., Burt D., See C.G., Elgar G., Sherrington R.,
RA Rogaev E.I., St George-Hyslop P.H., Brenner S., Dykes C.W.;
RT "Conservation of synteny between the genome of the pufferfish (Fugu
RT rubripes) and the region on human chromosome 14 (14q24.3) associated
RT with familial Alzheimer disease (AD3 locus).";
RL Proc. Natl. Acad. Sci. U.S.A. 93:1366-1369(1996).
CC -!- FUNCTION: NUCLEAR PHOSPHOPROTEIN WHICH FORMS A TIGHT BUT NON-
CC COVALENTLY LINKED COMPLEX WITH THE C-JUN/AP-1 TRANSCRIPTION
CC FACTOR. C-FOS HAS A CRITICAL FUNCTION IN REGULATING THE
CC DEVELOPMENT OF CELLS DESTINED TO FORM AND MAINTAIN THE SKELETON.
CC IT IS THOUGHT TO HAVE AN IMPORTANT ROLE IN SIGNAL TRANSDUCTION.
CC CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).
CC -!- SUBUNIT: HETERODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY. FOS SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U40757; AAC59778.1; -
CC HSSP; P01100; 1FOS.
CC DR TRANSFAC; T02205; -
CC DR PFAM; PF00170; BZIP; 1
CC DR PRINTS; PR00042; LEUZIPPROS.
CC DR PROSITE; PS00036; BZIP_BASIC; 1.
CC KW Proto-oncogene; Nuclear protein; Phosphorylation; DNA-binding.
CC FT DNA_BIND 123 144 BASIC MOTIF.
CC FT DOMAIN 149 177 LEUCINE-ZIPPER.
CC SQ SEQUENCE 376 AA; 40826 MW; BFC28534431DB491 CRC64;

Query Match 7.2%; Score 138.5; DB 1; Length 376;
Best Local Similarity 27.0%; Pred. No. 0.11;
Matches 53; Conservative 36; Mismatches 76; Indels 31; Gaps 7;

Qy 198 FVNFN-----DSAVG-----SIGGAELLGSLVDDVESTISFGSPSPETSQ 241
Db | : : : : : | : : : : : | : : : : : | : : : : : | : :
3 FTSFNAECDSSRCSASPVGNLYYPSPAGSYSSMGSPQSQDFTDLTAS----SASFPT 58
Qy 242 SSIIESSPELYKVIS--TSSIDASKRFPYSRSSKSKOSVKTSDA-KAPKTRTPAQVPV 298
Db | : : : : : | : : : : : | : : : : : | : : : : : | : :
59 VTAISTSPDLQWVPLVSSVAPSHRAHPYSPSPYKRTVMRSASAKHGRSRVEQTTP 118
Qy 299 EHVIMEHLDDKDRKLQNKNAIRYRMKKGEAQGIKEOELEINTKLTKVDDLORE 358
Db | : : : : : | : : : : : | : : : : : | : : : : : | : :
119 EEE-----EKKRIRERNKQAAKCRNRRLTDTLQAEITDLEKSSQLNDIANLKE 173
Qy 359 ---IKYMKNLMDVCK 371
Db | : : : : : | : : : : : | : : : : : | : : : : : | : :
174 KERLEFILAHHQPICK 189

RESULT 5
ATF2_CHICK
ID ATF2_CHICK STANDARD; PRT; 487 AA.
AC O93602;
DT 15-FEB-2000 (Rel. 39, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE CYCLIC-AMP-DEPENDENT TRANSCRIPTION FACTOR ATF-2 (ACTIVATING
DE TRANSCRIPTION FACTOR 2).

```

GN ATF2.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Huglier S., Baguet J., Perez S., van Dam H., Castellazzi M.;
 RT "Transcription factor ATF2 cooperates with vJun to promote growth-
 factor independent proliferation in vitro and tumor formation in
 vivo";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: A TRANSCRIPTIONAL ACTIVATOR, PROBABLY CONSTITUTIVE,
 WHICH BINDS TO THE CAMP RESPONSE REGION (CRE) (CONSENSUS:
 5'GTGACGT(A/C)(A/G)-3'); A SEQUENCE PRESENT IN MANY VIRAL AND
 CC CELLULAR PROMOTERS (BY SIMILARITY).
 CC -1- SUBUNIT: BINDS DNA AS A DIMER.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY. ATF SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Y17724; CAA76838.1; -;
 DR HSP; P08047; ISP2.
 DR PFAM; PF00170; bzip; 1.
 DR PFAM; PF00096; zf-C2H2; 1.
 DR PROSITE; PS00036; ZINC_FINGER_C2H2; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2; 1.
 KW Transcription regulation; DNA-binding; Activator; Nuclear protein;
 KW Zinc-finger; Metal-binding.
 FT ZN_FING 7 31 C2H2-TYPE.
 FT DOMAIN 333 356 BASIC MOTIF.
 FT DOMAIN 362 390 LEUCINE-ZIPPER.
 FT SEQUENCE 487 AA; 52406 MW; 41F42734D9C6A146 CRC64;
 SQ SEQUENCE 487 AA; 52406 MW; 41F42734D9C6A146 CRC64;
 Query Match 7.18; Score 137; DB 1; Length 487;
 Best Local Similarity 21.38; Pred. No. 0.19;
 Matches 84; Conservative 60; Mismatches 163; Indels 88; Gaps 15;
 QY 52 TLKLEPFEED-VLCAEWESDLSGLF--DALGDNHERLHPFESNLLEFTSLTPDDST 107
 DB 34 TLKFGPARNDSSVIVAD--QTPTTRFLKNCBEVGLFNLASPFEN---EFKASEDDIKK 88
 QY 108 VSKDILSSTLQFPQPNIPLYASHGAEDFSAETEFENHLSP-PDSPEQVAPVINLE--- 163
 DB 89 MPLDL-----SPLATPIIRNKIEPSVVEITHQDSPLPHPESTINDEKVSLOQTA 139
 QY 164 -----PVELTASHMTVISPDLGLGMEELASELFTTDLDFVNFDSAVGSIIGGAEL 215
 DB 140 QPTSTIVRPASLPVNPVLLTSSDSSVITQQAIPSTSTVITQATPSSNRPIVPVPGPEPL 199
 QY 216 L-----GSPLSVDD-----VESTISFGSPSPETSQS----- 242
 DB 200 LLHLPNGQTMVPAIPASITNSNVHVPAAVPLVRVMTVPSPGIPGPPSPQVQSEAKLR 259
 QY 243 ---SIIESSPELYKVIYSTSIDASK-----RFSYRSKSKQSVKTSDAKAPKTRT 292
 DB 260 LKAALTQHQVQV-----TNGDTAKHPSGLVTRQSEEPQPSLQQAATSTYTPASPAPQ 313
 QY 293 PAQVVP-----EHVMEHLDDKKRKLQ-KNNAAIYRKMKGGAQIGEEQEELN 345
 DB 314 PTQQTPTNGRRRAANEDPDDEKRKFLERNRAAASRCQKRKVVWVQSLKKAEDLSLN 373
 QY 346 TLKTKVDLDLQREIKYMKNLW--EDVCCKAGIOLK 378
 DB 374 GOLONEVTLRNEVAQLQLLLAHKCPVTAMQKK 408

RESULT 6
 ATF2_HUMAN
 ID ATF2_HUMAN STANDARD; PRT; 487 AA.
 AC P15336;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE CYCLIC-AMP-DEPENDENT TRANSCRIPTION FACTOR ATF-2 (ACTIVATING
 DE TRANSCRIPTION FACTOR 2) (CAMP RESPONSE ELEMENT BINDING PROTEIN CRE-
 DE BPL1) (HB16).
 DE BPL1 (HB16).
 GN ATF2 OR CREB2 OR CREBPL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Maekawa T., Sakura H., Kanei-Ishii C., Sudo T., Yoshimura T.,
 RT Fujisawa J.I., Yoshida M., Ishii S.;
 RT "Leucine zipper structure of the protein CRE-BPL binding to the
 RT cyclic AMP response element in brain.";
 RL EMBO J. 8:2023-2028(1989).
 CC [2]
 CC SEQUENCE OF 193-487 FROM N.A.
 CC MEDLINE; 90705810.
 CC "A CDNA for a human cyclic AMP response element-binding protein which
 CC is distinct from CREB and expressed preferentially in brain.";
 CC Mol. Cell. Biol. 10:1347-1357(1990).
 CC -1- FUNCTION: THIS PROTEIN BINDS THE CAMP RESPONSE ELEMENT (CRE)
 CC VIRAL AND CELLULAR PROMOTERS.
 CC -1- SUBUNIT: BINDS DNA AS A DIMER.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- ALTERNATIVE PRODUCTS: MXBP AND CRE-BPL MAY BE MEMBERS OF A FAMILY
 CC OF MXBP/CRE-BP PROTEINS GENERATED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: ABUNDANT EXPRESSION SEEN IN THE BRAIN.
 CC -1- PTM: PHOSPHORYLATION OF THR-69 AND THR-71 BY MAPK14 CAUSES
 CC INCREASED TRANSCRIPTIONAL ACTIVITY. ALSO PHOSPHORYLATED AND
 CC ACTIVATED BY JNK.
 CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY. ATF SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X15875; CAA33886.1; -;
 DR EMBL; M31630; AAA35951.1; -;
 DR PIR; S05380; S05380.
 DR HSP; P08047; ISP2.
 DR TRANSFAC; T00167; -;
 DR TM; 123811; -;
 DR PFAM; PF00170; bzip; 1.
 DR PFAM; PF00096; zf-C2H2; 1.
 DR PROSITE; PS00036; ZINC_FINGER_C2H2; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2; 1.
 KW Transcription regulation; DNA-binding; Activator; Phosphorylation;
 KW Nuclear protein; Alternative splicing; Zinc-finger; Metal-binding.
 FT ZN_FING 7 31 C2H2-TYPE.
 FT DOMAIN 333 356 BASIC MOTIF.
 FT DOMAIN 362 390 LEUCINE-ZIPPER.
 FT MOD_RES 51 51 PHOSPHORYLATION (BY MAPK14).
 FT MOD_RES 53 53 PHOSPHORYLATION (BY MAPK14).
 FT SEQUENCE 487 AA; 52249 MW; EEE4D17FE8DB7CD7 CRC64;
 SQ SEQUENCE 487 AA; 52249 MW; EEE4D17FE8DB7CD7 CRC64;


```
ID YDC3_SCHPO STANDARD; PRT; 330 AA.
AC Q10424;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HPOTHEICAL 36.2 KDA PROTEIN C25G10.03 IN CHROMOSOME 1.
GS SPAC25G10.03.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC Schizosaccharomycetaceae; Schizosaccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA McLean J., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -1- SIMILARITY: TO OTHER BZIP PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z70691; CAA94632.1; -.
CC HSP; P03059; 2DGC.
CC PROSITE; PS00036; BZIP_BASIC; 1.
KW Hypothetical protein; Transcription regulation; DNA-binding;
FT DNA_BIND 270 288 BASIC MOTIF (BY SIMILARITY).
FT SEQUENCE -330 AA; 36189 MW; A681434C779DF960 CRC64;
SQ SEQUENCE 330 AA; 36189 MW; A681434C779DF960 CRC64;

Query Match 7.0%; Score 134.5; DB 1; Length 330;
Best Local Similarity 22.0%; Pred. No. 0.16;
Matches 72; Conservative 54; Mismatches 130; Indels 71; Gaps 10;

QY 95 LEFTSLITPDSTVSKDLSLTQFPQPVNIPL----YASHGAEDFSAEFEFENHLSPP 150
DB 14 LKFDVDPVSD--FSKDDLABQLNFTNPFDLDEPSMLSEGYGVGVSQSGSSNKKQ 71

QY 151 D-----SPEQAPVINLEPVELTASHMTVISPDGLGGMELASESLTFELD----FVN 200
DB 72 EKNVQOQNPKEIKISTLQQVKEEVNTESAPLNATGNFSSANPASIDLAYLDLKLTLPD 131

QY 201 FNDASVGSIGGAELLGSLV-----DDVESTISFGPSS-----236
DB 132 HSKETQKTSQRLEFPQKSVASASKDNVSSSILQGSASSKLLPDQSARQHOVLGQT 191

QY 237 --PETSQSIIESSPELVK-----IST-SSIDASKRFSYRSKSKSQS--V 279
DB 192 AIPTEASSSSINNTPLQAVSSFDQAFNPLTFASPDLASVSSLSYKGAQSPNA 251

QY 280 KTSDAKAPRTKTPAQVPEHIVHELDKDKKLLKLNKAAIRYMRKKGGAQGIKGEQ 339
DB 252 NSKRTKATSAIRTAEE-----EDKRRRTAASAREFRICKLKEQQLERTAK 297

QY 340 ELEELNTKLTQVDDQLQREIKYMKNL 366
DB 298 ELTEKVAILETRVRELEMENNWLKGLI 324

RESULT 9
ATP2_RAT
ID ATP2_RAT STANDARD; PRT; 487 AA.
AC Q00969; Q62870;
DT 01-APR-1993 (Rel. 25, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE CYCLIC-AMP-DEPENDENT TRANSCRIPTION FACTOR ATF-2 (ACTIVATING
DE TRANSCRIPTION FACTOR 2) (CAMP RESPONSE ELEMENT BINDING PROTEIN CRE-
```

```
DE BP1).
GN ATF2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC SYRAIN-SPRAGUE-DAWLEY; TISSUE-BRAIN;
RA Muramatsu S.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE-BRAIN;
RA Kageyama R., Sasai Y., Nakanishi S.;
RL "Molecular characterization of transcription factors that bind to the
RT CAMP responsive region of the substance P precursor gene. CDNA
RT cloning of a novel C/EBP-related factor."
CC J. Biol. Chem. 266:15525-15531(1991).
CC -1- FUNCTION: A TRANSCRIPTIONAL ACTIVATOR, PROBABLY CONSTITUTIVE,
CC WHICH BINDS TO THE CAMP RESPONSIVE REGION (CRE) (CONSENSUS:
CC 5'GTGACGT(A/C)(A/G)-3'); A SEQUENCE PRESENT IN MANY VIRAL AND
CC CELLULAR PROMOTERS.
CC -1- SUBUNIT: BINDS DNA AS A DIMER.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: ISOFORM 1 (SHOWN HERE) AND
CC ISOFORM 2: ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY. ATF SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U38938; AAA93263.1; -.
CC EMBL; M65148; AAA42013.1; -.
CC PIR; A39429; A39429.
CC HSP; P08047; ISP2.
CC TRANSFAC; T01382; -.
CC PFAM; PF00170; bzip; 1.
CC PFAM; PF00096; zf-C2H2; 1.
CC PROSITE; PS00036; BZIP_BASIC; 1.
CC PROSITE; PS00036; ZINC_FINGER_C2H2; 1.
CC PROSITE; PS00028; ZINC_FINGER_C2H2; 1.
KW Transcription regulation; DNA-binding; Activator; Phosphorylation;
KW Nuclear protein; Alternative splicing; Zinc-finger; Metal-binding.
FT ZN_FING 7 31
FT C2H2-TYPE 31
FT DNA_BIND 333 356 BASIC MOTIF.
FT DOMAIN 362 390 LEUCINE-ZIPPER.
FT MOD_RES 51 51 PHOSPHORYLATION (BY MAPK14) (BY
FT MOD_RES 53 53 PHOSPHORYLATION (BY MAPK14) (BY
FT MOD_RES 53 53 PHOSPHORYLATION (BY MAPK14) (BY
FT VARSPLIC 132 229 MISSING (IN ISOFORM 2).
FT SEQUENCE 487 AA; 52286 MW; 4ED95B106DF5F9EE CRC64;

Query Match 7.0%; Score 134; DB 1; Length 487;
Best Local Similarity 21.9%; Pred. No. 0.29;
Matches 86; Conservative 58; Mismatches 166; Indels 82; Gaps 16;

QY 52 TLKLEPFEED-VLGAEMWESSDLGSLF---DALGDNHERLHPESNLLETSLITPDST 107
DB 34 TLKFGPARNDSVIVAD--QTPTPTFLKNCSEVGLFNELASPFEN---EFKASEDDIKK 88

QY 108 VSKDILSSTLQFPQPVNIPLYASHGAEDFSAEFEFENHLSPP-PDS---PEQVAPVIN-- 161
DB 89 MPDL-----SPLATIIISKTEEPSVVTTHQDSPLPHPPESTYNDKEIPLAQTA 139

QY 162 -----LEPVELTASHMTVISPDGLGGMELASESLTFELDFVNFNDSAVSIGGAEL 215
```

```
Db 140 QPSTAIVRPAQLQVNVLLTSSDSSVLIQQAQVPSPTSSVTITQAPSSNRPIVVPGRPL 199
QY 216 L-----GSLSVDD-----VESTISFGSPSPETQSIIIES 247
Db 200 LHLPLNGQTPVAIPAISTSSNVHVPAAVPLVRPVTMVPSPGIPGSPQVQS-----EA 256
QY 248 SPFLYKVIS-----TSSIDASKRSPYSRSKSKQSVKTSDAKAPKTRTPAQP---VP 298
Db 257 KWLRLAALTOQHPPTVNGDVTYKGGSLVRAQSESRQSLQQAQYATSTETPASPAHTP 316
QY 299 E-----HVIMEHLDKDKRKQLQ--NKNAIRYMRKKGGAQKIGKEEQEELBELNTRL 348
Db 317 QFQNTSGRRRAANEDPDEKRKRFELERNRAAASRCQRKRYVWVQSLKKAEDLSSINGOL 376
QY 349 KTKVDLDLQREIKYMKNL--EDVCKAKGIQLK 378
Db 377 QSEVTLRNEVAQLKOLLAHKDCPVTAMQK 408

RESULT 10
SKOL_YEAST STANDARD; PRT; 647 AA.
AC Q02100;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE CRE-BINDING BZIP PROTEIN SKO1.
GN SKO1 OR ACR1 OR YNL167C OR NI702.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-W303-1A;
RA MEDLINE; 93065191.
RA Nehlin J.O., Carlberg M., Ronne H.;
RT "Yeast SKO1 gene encodes a bZIP protein that binds to the CRE motif
and acts as a repressor of transcription.";
RL Nucleic Acids Res. 20:5271-5278(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93078739.
RA Vincent A.C., Struhl K.;
RT "ACR1, a yeast ATF/CREB repressor.";
RL Mol. Cell. Biol. 12:5394-5405(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / FY1679;
RX MEDLINE; 96287653.
RA Nasr F., Becam A.-M., Herbert C.J.;
RT "The sequence of 36.8 kb from the left arm of chromosome XIV reveals
24 complete open reading frames: 18 correspond to new genes, one of
which encodes a protein similar to the human myotonic dystrophy
kinase.";
RL Yeast 12:169-175(1996).
CC -!- FUNCTION: BINDS TO THE CRE MOTIF 5'-TGACGTCA-3' AND ACTS AS A
REPRESSOR OF TRANSCRIPTION OF THE SUC2 GENE AND MOST PROBABLY
OTHER GENES.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: TO OTHER BZIP PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X67875; CAA48074.1; -
DR EMBL; S49588; AAB24288.1; -
DR EMBL; X92317; CAA63272.1; -
DR EMBL; Z71443; CAA96054.1; -
```

```
DR PIR; S26386; S26386.
DR PIR; A45028; A45028.
DR HSP; P05412; 1FOS.
DR TRANSFAC; T01306; -.
DR SGD; L0001909; SKO1.
DR PIR; P00170; bZIP; 1.
DR PROSITE; PS00036; BZIP_BASIC; FALSE_NEG.
KW Transcription regulation; Repressor; DNA-binding; Nuclear protein.
FT DNA_BIND 431 456 BASIC MOTIF.
FT DOMAIN 457 478 LEUCINE-ZIPPER.
SQ SEQUENCE 647 AA; 70192 MW; 3E0B8C72A6CE14B CRC64;

Query Match 7.0%; Score 134; DB 1; Length 647;
Best Local Similarity 21.1%; Pred. No. 0.43;
Matches 75; Conservative 60; Mismatches 132; Indels 88; Gaps 14;

QY 45 SLOQAGATLKLEPFEEDVLGAEMESSDLGSLDALGDHNRHLHPESNLLEFTSLITPD 104
Db 169 SSSPGTSLIRPNSNV-----TTSNG---NFPTNDSQMPGFLLLSK--SGLTPN 217
QY 105 DSTVSKDILSSTLQFPTQPVNIPLYASHGAEDFSAETEFENHLSPDPSPEQ-----VAP 158
Db 218 ESNIRTLGTPGIL---TQSYNYPVLPISINKNTITGSKNVKSVTVNGSIENHPVNMHP 274
QY 159 VINLPVELTASHMVTVISPDGLLGGMELASELSLTTELDVFNPNDSAVSGISGAELLGS 218
Db 275 TVNGTPLTPGLSLLNLPSTGVLANPVPFKSTPTT-----NTTDTG----- 314
QY 219 PLSVDDVESTISFGSPSPETQSIIIE--SPELYKVIS-----TSSIDASKR 265
Db 315 -----VNNSISNSN-FSPNTSTKAAVKMDNPAEINAEISAHNIKENENLTQIENDDO 367
QY 266 FSPYSRSKSKQSVKTSDAKAPRK-----TRTPAQ-----PVPEHVIMEHLDK 308
Db 368 FNNKTKRKRMRMSSSTSSKSKNSISRKNSAVTAPAKDDVENNKTSNNVTLDNEE 427
QY 309 KDRKKLQ-----NKNAIRYMRKKGGAQKIGKEBELEELNKLTKYVDLQREI 359
Db 428 QERKRKEFLERNVAASKPRKRKEYIKKIENDLQFYE-----SEYDULQTVI 475

RESULT 11
ATFA_HUMAN STANDARD; PRT; 483 AA.
AC P17544;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE TRANSCRIPTION FACTOR ATF-A AND ATF-A-DELTA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90301459.
RA Gaire M., Chatton B., Kedinger C.;
RT "Isolation and characterization of two novel, closely related ATF
cDNA clones from HeLa cells.";
RL Nucleic Acids Res. 18:3467-3473(1990).
CC -!- FUNCTION: THIS PROTEIN BINDS THE CAMP RESPONSE ELEMENT (CRE)
VIRAL AND CELLULAR PROMOTERS.
CC -!- SUBUNIT: BINDS DNA AS A DIMER.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- ALTERNATIVE PRODUCTS: ATF-A AND ATF-A-DELTA, WHICH ARE IDENTICAL,
EXCEPT FOR ADDITIONAL 21 AMINO ACIDS IN ATF-A (RESIDUES 114-134)
MAY BE DERIVED FROM A COMMON PRIMARY TRANSCRIPT BY ALTERNATIVE
SPLICING EVENTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
```

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; X52943; CAA37118.1; -
 DR PIR; S12741; S12741.
 DR HSP; P03069; 2ZTA.
 DR TRANSFAC; T00052; -
 DR TRANSFAC; T00053; -
 DR PFAM; PF00170; bZIP; 1.
 DR PFAM; PF00096; zf-C2H2; 1.
 DR PROSITE; PS00036; BZIP_BASIC; 1.
 DR Transcription regulation; DNA-binding; Activator; Nuclear protein;
 KW Phosphorylation; Alternative splicing.
 FT DNA_BIND 331 354 BASIC MOTIF.
 FT DOMAIN 360 388 LEUCINE-ZIPPER.
 FT VARSPPLIC 114 134 MISSING (IN ISOFORM ATF-A-DELTA).
 SQ SEQUENCE 483 AA; 51756 MW; 07080BC24FED635B CRC64;

Query Match 6.9%; Score 133; DB 1; Length 483;
 Best Local Similarity 24.08; Pred. No. 0.33;
 Matches 97; Conservative 60; Mismatches 138; Indels 110; Gaps 23;

QY 52 TLKLEPFEED-VLGAENWESSDLGSFL---DALGDNHERLHPFESNLLEFTSLITPDST 107

DB 34 TLKFGPARTDSVIAD-QPTPTFRFLKNEEVGLFNLASSFEH---EPKKADEDEKK 88

QY 108 VSKDILSTTQFPQPNIPLYASHGAEDFSAETEFENHLSPPDSP-----EQV 156

DB 89 AAAGPLDMSL--PSTP-DIKI-----KEEPVEVDSPDPSPASSPCPPPLKEVEV 136

QY 157 APVINLPVELTASHMTVISPDGI---LG---GMELASLFTLTETDFNFNDSAVGSI 209

DB 137 TP---APVLISPTPTIVRPGSLPLHLGYDPLHPTLPSTSVITQPPSN---RQMGSP 189

QY 210 GGABELL-----GSLPSVDVDESTISFSGPSS-----PETSQSII 245

DB 190 TGSPLPLVHLANGQTPVLPQPPV---QMPSVISLARVPSVMPNIPGPPVNSSGSI- 245

QY 246 ESSPELYKVIS-----TSSID-----ASKRSPYSRSKSKOSVTS 283

DB 246 --SPSGHPISPEAKMRLKATLTHOVSSINGCGMVVGTAFTMTV--ARPEQSQILLIHPD 301

QY 284 AKAPRKR-TPAOPVP-----EHVIMEHLDKDRKKLO-NKNAIYRMYKKKGAEAGCIK 335

DB 302 APSPAQPVPAQPTPTSGTGGRRRTVDEPDERRQRFLERNRAAASCRKRKLWSSLE 361

QY 336 GEEQEELTNTKLTKYDDLLQREIKYMKNLN---EDVCARKGIQIK 378

DB 362 KKAELTSQNLQLSNEVTLRNEVAQLKQLLAHKDCPVTALQKK 406

RESULT 12
 ATF3_HUMAN
 ID ATF3_HUMAN STANDARD; PRT; 181 AA.
 AC P18847;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE CYCLIC-AMP-DEPENDENT TRANSCRIPTION FACTOR ATF-3 (ACTIVATING
 DE TRANSCRIPTION FACTOR 3).
 GN ATF3.
 GN ATFL3.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE; 94253175.
 RA Chen B.P.C., Liang G., Whelan J., Hai T.;
 RT "ATF3 and ATF3 delta zip. Transcriptional repression versus
 RT activation by alternatively spliced isoforms."

RL J. Biol. Chem. 269:15819-15826(1994).
 RN [2]
 RP SEQUENCE OF 25-181 FROM N.A.
 RX MEDLINE; 90185187.
 RA Hai T., Liu F., Coukos W.J., Green M.R.;
 RT "Transcription factor ATF cDNA clones: an extensive family of leucine
 RT zipper proteins able to selectively form DNA-binding heterodimers.";
 RL Genes Dev. 3:2083-2090(1989).
 RN [3]
 RP ERRATUM.

RA Hai T., Liu F., Coukos W.J., Green M.R.;
 RL Genes Dev. 4:682-682(1990).
 CC -1- FUNCTION: THIS PROTEIN BINDS THE CAMP RESPONSE ELEMENT (CRE)
 CC (CONSENSUS: 5'-GTGACGT[AC][AG]-3'). A SEQUENCE PRESENT IN MANY
 CC VIRAL AND CELLULAR PROMOTERS. REPRESSSES TRANSCRIPTION FROM
 CC PROMOTERS WITH ATF SITES. IT MAY REPRESS TRANSCRIPTION BY
 CC STABILIZING THE BINDING OF INHIBITORY CO-FACTORS AT THE PROMOTER.
 CC AN ALTERNATIVE SPLICED FORM (ATF3-DELTA-ZIP) ACTIVATES
 CC TRANSCRIPTION PRESUMABLY BY SEQUESTERING INHIBITORY CO-FACTORS
 CC AWAY FROM THE PROMOTERS.

CC -1- SUBUNIT: BINDS DNA AS A HOMODIMER OR A HETERODIMER.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.

CC -1- ALTERNATIVE PRODUCTS: TWO FORMS ARE PRODUCED BY ALTERNATIVE
 CC SPLICING. THE SHORT FORM (ATF3-DELTA-ZIP) LACK THE LEUCINE ZIPPER
 CC DOMAIN.

CC -1- SIMILARITY: TO OTHER BZIP PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; L19871; AAA20506.1; -
 DR PIR; C34223; C34223.
 DR HSP; P01100; 1FOS.
 DR TRANSFAC; T01313; -
 DR TRANSFAC; T01325; -
 DR MIM; 603148; -
 DR PFAM; PF00170; bZIP; 1.
 DR PROSITE; PS00036; BZIP_BASIC; 1.
 DR Transcription regulation; DNA-binding; Repressor; Nuclear protein;
 KW Multigene family; Alternative splicing.
 FT DNA_BIND 88 110 BASIC MOTIF.
 FT DOMAIN 114 142 LEUCINE-ZIPPER.
 FT VARSPPLIC 116 181

FT CONFLICT 132 132 I -> L (IN REF. 2).
 FT SEQUENCE 181 AA; 20575 MW; EC5D8F065EEE2D9C CRC64;

Query Match 6.6%; Score 127; DB 1; Length 181;
 Best Local Similarity 29.3%; Pred. No. 0.21;
 Matches 44; Conservative 31; Mismatches 55; Indels 20; Gaps 6;

QY 235 SSPETSOSIIIES-SPELYKVIS-----TSSIDASKRSPYS-----RSSKSKOSVTS 283

DB 10 SASEVSASAIIVCLSPGSLVDFEDFANLTPFVKEELFAIQNKHLCHRMSSALESVTS 69

QY 284 AKAPRTRTPAOPVPEHVIMEHLDKDRKKLONKNAIRYRMYKKKGAEAGCIK 343

DB 70 -RPLGVISITAEVAPED-----ERKARRRNRKIAAOKRNRKKECTCQSEKLES 123

QY 344 LNTKTKTKVDDLQRE---IKYMKNLMDVC 370

DB 124 VNAELKAEIELKNEKHLYMLNLHPTC 153

RESULT 13

PANI_RAT

ID AC PAN1_RAT STANDARD; PRT; 638 AA.
DT P21676;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE TRANSCRIPTIONAL REGULATORY PROTEIN PAN-1 (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90346284.
RA Nelson C., Shen L.-P., Meister A., Fodor E., Rutter W.J.:
RT "Pan: a transcriptional regulator that binds chymotrypsin, insulin,
and AP-4 enhancer motifs.";
RL GENES Dev. 4:1035-1043(1990).
CC -1- FUNCTION: TRANSCRIPTIONAL REGULATOR THAT BINDS TO THE CONSENSUS
SEQUENCE CAC/GCTGT/C PRESENT, IN THE CHYMOTRYPSIN, INSULIN, AP-4,
AND SEVERAL OTHER GENE ENHANCER MOTIFS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: PAN-1 AND PAN-2 MAY BE GENERATED BY
ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
TRANSCRIPTION FACTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X62323; CAA44199.1; .
DR PIR; B35816; B35816.
DR HSP; P10085; 1MDY.
DR TRANSFAC; T00674; .
DR PFAM; PF00010; HLH; 1.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
KW Transcription regulation; DNA-binding; Nuclear protein; Repeat;
KW Alternative splicing.
FT NON_TER 1 1
FT REPEAT 180 184 S-S-S-C-D.
FT REPEAT 314 318 S-S-S-G-D.
FT DOMAIN 377 412 LEUCINE-ZIPPER (POTENTIAL).
FT DNA_BIND 533 587 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
SQ SEQUENCE 638 AA; 66350 MW; 74489490C97D2BE5 CRC64;

Query Match 6.6%; Score 127; DB 1; Length 638;
Best Local Similarity 21.7%; Pred. No. 1.1;
Matches 96; Conservative 54; Mismatches 146; Indels 146; Gaps 23;

QY 4 DLWSEDFQ-----LARENGLEMPVQV-----TDGQFGDLKSTR-----HGGD----- 41
DB 219 ELWPPGQVGFGLMGDSAPLPIAPGSSSVSSGAFGLQGDQRMGVLHGFSEVNGTLPA 278
QY 42 -ESLSLOP--QGATLKLEPPE--EDVLGAEMWESSDLGFLDALGDNHRHLHPFESNLL 95
DB 279 VSSFSAPGTYSGTSGHTPPVSGADSLGTRGTASSG---DALGK----- 322
QY 96 EFTSLTPDDSTVSKDILSSTLQFPPTQVNPILYASHGAEDFSAETEPENHLSPDPSPEQ 155
DB 323 ALASIYSPD-----HSSNNFSPSPS-----TPVGGSP-Q 349
QY 156 VAPVNLPEVELTASHMTVISP--DGLLGGMELASESLTFTLDFVFNDSAV-----GS 208
DB 350 GLPGTSQMP---RAGAPSAISPNDAGLGLSKMEDRL-----DEAIHVLRSR 395
QY 209 IGGAEELGSLPSVDDVESTISFSGP-----SSPE---TSQSSII----- 245
DB 396 VGTASELHG--LLPGHSLTLTTSFAGPMSLGRHAGLVSGSHPEDGLTSCALLHNHSLP 453

QY 246 -----BSSPELYKVISSIDASKRPSRSPSSKSKQSVKT-----SDAKAP 287
DB 454 SOPSPLDLSQRPPDSFSGLRAGVTAGA--SEIKREEKEDEEVTSTVADAEDKKDLKVP 511
QY 288 RKTRPAQVPPEHVIMEHLDKDRKKLQNKNAAIYRMKKKGGAQIGKEGEQEELELNK 347
DB 512 R-TRTSS--TDEVLSELEKDLDRERRMANNARVRVDINEAFRLGRMCOLHLKSDK 568
QY 348 LKTKVDDLQREIKYMKNLMDV 369
DB 569 AQTKLLILQQAQVQVILGLEQQV 590

RESULT 14
MAPA_RAT
ID MAPA_RAT STANDARD; PRT; 2774 AA.
AC P34926;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE MICROTUBULE-ASSOCIATED PROTEIN 1A [CONTAINS: MAP1 LIGHT CHAIN LC2].
GN MAP1A.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-BRAIN;
RX MEDLINE; 92355629.
RA Langkopf A., Hammarback J.A., Mueller R., Vallee R.B., Garner C.C.;
RT "Microtubule-associated proteins 1A and LC2. Two proteins encoded in
one messenger RNA.";
RL J. Biol. Chem. 267:16561-16566(1992).
CC -1- FUNCTION: STRUCTURAL PROTEIN INVOLVED IN THE FILAMENTOUS
CROSS-BRIDGING BETWEEN MICROTUBULES AND OTHER SKELETAL ELEMENTS.
CC -1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN-ASSOCIATE
WITH MAP1A AND MAP1B PROTEINS.
CC -1- TISSUE SPECIFICITY: BRAIN, HEART AND MUSCLE.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED LATE DURING NEURONAL DEVELOPMENT
APPEARING WHEN AXONS AND DENDRITES BEGIN TO SOLIDIFY AND STABILIZE
THEIR MORPHOLOGY.
CC -1- DOMAIN: THE BASIC REGION CONTAINING THE REPEATS MAY BE RESPONSIBLE
FOR THE BINDING OF MAP1A TO MICROTUBULES.
CC -1- PTM: VARIOUS SERINE RESIDUES MAY BE PHOSPHORYLATED BY CAMP KINASE.
CC -1- PTM: LC2 IS COEXPRESSED WITH MAP1A. IT IS A POLYPEPTIDE GENERATED
FROM MAP1A BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
BOTH MAP1A AND MAP1B.
CC -1- SIMILARITY: TO MAP1B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M83196; AAB48069.1; .
DR PIR; A43359; A43359.
KW Microtubules; Repeat; Phosphorylation.
FT CHAIN 72465 2774 MAP1 LIGHT CHAIN LC2.
FT DOMAIN 309 496 LYS-RICH (BASIC).
FT REPEAT 336 541 11 X 3 AA REPEATS OF K-R-[DE].
FT REPEAT 336 338 1.
FT REPEAT 415 417 2.
FT REPEAT 420 422 3.
FT REPEAT 424 426 4.
FT REPEAT 427 429 5.
FT REPEAT 431 433 6.
FT REPEAT 436 438 7.
FT REPEAT 440 442 8.
FT REPEAT 444 446 9.
FT REPEAT 449 451 10.

```
FT REPEAT 539 541 11.
SQ SEQUENCE 2774 AA; 299526 MW; 3DEF74427BA9D7D7 CRC64;

Query Match 6.58; Score 124.5; DB 1; Length 2774;
Best Local Similarity 22.3%; Pred. No. 11;
Matches 91; Conservative 50; Mismatches 142; Indels 125; Gaps 20;

QY 30 GDLKSTSRHG-GDESLSLQV-----GATLKLEPEEDVLGAENWES-SDLGSGF 76
Db 1045 GPEGTSEEGKGPTRSPQADMPVSTAGGTCTIQLLPQDKAIVFETGEAGSNLGG 1104
QY 77 L---DALGDNHERLHPESNLLEFT-SLITPDD-----STVSKDILSSTLQ--FTQP 123
Db 1105 TLPCEVSTSEATEPQKDEVLRTDQSLSPEDAESLSVSVSPD---TTKQEAATRSP 1161
QY 124 VNI-----PLYA-----SHGAEDFAETEE---NHLSPDP-----S 152
Db 1162 CSLKEQPHKDLMPVSPEDTQSLSFSESPSKETSLDISKQLSPESLGTQFGELNLG 1221
QY 153 PEQVAPVI-----NLEPVLTASHMTVISPDLGLGMELASESLFTFLDFVNFDSA 205
Db 1222 KEERGVPWKAEDDSCHLAPVSIPEPHRATVSP-----STDETP 1259
QY 206 VGSIGGAELGLSPSLVDVESTISFSGPSPSPETSQSSIIESSPELYKVIISTSIDASKR 265
Db 1260 AGTLPGG-SFSHSALSVDKHSPEITGCGHFMTSDSLTKSP----- 1303
QY 266 FSPYSRSKSKQSVKTS-DAKAPKRTTP-----AQVPEHVIHMLD-----KKD 310
Db 1304 ----SLSSPAMEDLAVWECKAPGKEREPELKTROQKGOILPEKVAVVEQDLIIHQK 1359
QY 311 RKLQNKAAIRYRMKKGAQGIKGEQELEELNLTNLTNKTVDLDQRE 358
Db 1360 -GALDEENKFGROQDKTPQKGRDLDEKDTAAELDKGPEPKELDRE 1406

RESULT 15
ENPL_TORCA STANDARD; PRT; 721 AA.
AC PI4400;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JUN-1990 (Rel. 13, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE ELECTROMOTOR NEURON-ASSOCIATED PROTEIN 1 (FRAGMENT).
OS Torpedo californica (Pacific electric ray).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Neoselachii; Squalia; Hypnosqualia; Pristioraja;
OC Batoidea; Torpediniformes; Torpedinidae; Torpedo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90091742.
RA Ngsee J.K., Scheller R.H.;
RT "Isolation and characterization of two homologous cDNA clones from
RT Torpedo electrotomotor neurons. ";
RL DNA 8:555-561(1989).
CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH MEMBRANES OF INTRACELLULAR
CC ORGANELLES.
CC -1- MISCELLANEOUS: HAS A HIGHLY BASIC REGION WITH MANY COPIES OF THE
CC SEQUENCE KKE, REPEATED BUT NOT AT FIXED INTERVALS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; M30270; AAA49279.1;
DR PIR; A33319; A33319.
DR HSSP; P11387; 1A36.
KW Neurone; Repeat.
```

```
FT NON_TER 1 1
FT SIMILAR 233 582
FT 80% IDENTITY TO ELECTROMOTOR NEURON-
FT ASSOCIATED PROTEIN 2.
FT GLU/LYS-RICH (BASIC).
FT 11 X 3 AA APPROXIMATE REPEATS.
FT 1.
FT 2.
FT 3.
FT 4.
FT 5.
FT 6.
FT 7.
FT 8.
FT 9.
FT 10.
FT 11.
FT NON_TER 721 721
SQ SEQUENCE 721 AA; 81241 MW; B53E69A9F309CE00 CRC64;

Query Match 6.3%; Score 121; DB 1; Length 721;
Best Local Similarity 20.5%; Pred. No. 2.9;
Matches 88; Conservative 68; Mismatches 157; Indels 116; Gaps 19;

QY 38 HGGDESL-----LPOGATLKLPEEEDVLGAEMW-----SSDLGSLDALGDNHERLHP 89
Db 316 HNGADNLPGVNSLLQRIKIAELEEEQSGSTANSWMKNLISPELCVFLNVPENVNLQ 375
QY 90 -----FESNLLEFTSL-----ITPDD--STVSKDILSSTLQFTQPV-NIPLYASHGAE 135
Db 376 NFRVRRNTEETCTLTQVILKNCVKPEPLFTVG-NVIDPVILFQKMGVGRLEWILNSVK 434
QY 136 DFSAETEFENHLSPDPSPQVAPVINLEPVLTASHMTVIS----- 176
Db 435 GSKELQFFMQHWSNNKAKTGLILPNGKDAEISPPYLTSSISLWHPANPSEKIVRALF 494
QY 177 ----PD-GLLGMELASESLTFTFLDFVNFNDSAVGSGAEELGSLPS----- 221
Db 495 PGNAQVNIIDGLE-----KLKHLDFLKH-----PVTKKELIASSAAPTQKAKLKQ 542
QY 222 -VDDVESTISFGSPSPETSSQSSIIESSPELYKVIISTSIDASKRFPYSRSKSKSVK 280
Db 543 WSDSKESLSKNSRPSVGKGVKDKVKEETPELTPTAVSHQEAUNE-KPQKVEKKEKPVVK 601
QY 281 TSDAKAPKRTTTPAQVPE-----HVTMEHLDKDKKKLQNKNAAIRYRMKKK 328
Db 602 -----KERPTELOSKPEKDAKAKAAQKQLEKMKQDKELKSESKPKPLKEKIVKK 655
QY 329 GEAQGIKGEQELEELNLTNLT-----KTKVDDLDQREIKYMKNL 365
Db 656 -EVKAKPPEKKKKEKDKVESAKPDRKEKAVIKKEKVKKEKPKKEVKKDKV--KDV 712
QY 366 MEDVCKARG 374
Db 713 KKEV-KKKG 720

Search completed: August 8, 2000, 06:12:24
Job time: 13118 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2000, 01:43:29 ; Search time 39.91 Seconds
(without alignments)
658.422 Million cell updates/sec

Title: US-08-656-811a-1
Perfect score: 1928
Sequence: 1 MELDLWSEDFQLAREWGLEM.....KYMKNLMEDVCVAKGIQLKX 379

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 59334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_12:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mmc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1923	99.7	378	5 Q16946	Q16946 aplysia cal
2	248	12.9	381	11 Q61328	Q61328 mus musculus
3	246.5	12.8	354	13 Q9W610	Q9W610 gallus gall
4	188.5	9.8	182	4 Q9Y201	Q9Y201 homo sapien
5	162	8.4	84	11 Q70191	Q70191 mus musculus
6	148.5	7.7	488	5 Q23272	Q23272 caenorhabdi
7	143.5	7.4	208	5 Q22156	Q22156 caenorhabdi
8	137.5	7.1	494	4 Q13814	Q13814 homo sapien
9	137	7.1	358	11 Q64091	Q64091 mus musculus
10	137	7.1	456	11 Q64090	Q64090 mus musculus
11	137	7.1	487	13 Q93602	Q93602 gallus gall
12	135	7.0	486	13 Q91576	Q91576 xenopus lae
13	134	7.0	487	11 Q62870	Q62870 rattus norv
14	133	6.9	448	11 Q64089	Q64089 mus musculus
15	131	6.8	2390	11 Q08592	Q08592 rattus norv
16	129.5	6.7	457	5 Q19521	Q19521 caenorhabdi
17	127.5	6.6	1414	11 Q63330	Q63330 rattus norv
18	127	6.6	460	5 Q17801	Q17801 caenorhabdi
19	126.5	6.6	338	5 Q16949	Q16949 aplysia cal

20	126	6.5	699	11 Q35451	Q35451 mus musculus
21	126	6.5	938	4 Q13111	Q13111 homo sapien
22	125	6.5	507	11 Q92125	Q92125 mus musculus
23	124.5	6.5	1142	4 Q60732	Q60732 homo sapien
24	124.5	6.5	1142	4 Q75451	Q75451 homo sapien
25	123	6.4	329	4 Q13000	Q13000 homo sapien
26	123	6.4	613	5 Q9X258	Q9X258 drosophila
27	121	6.3	424	10 Q41109	Q41109 phaseolus v
28	120.5	6.2	780	10 Q04346	Q04346 arabidopsis
29	119.5	6.2	319	13 Q91639	Q91639 xenopus lae
30	119	6.2	341	3 Q12191	Q12191 saccharomyc
31	119	6.2	700	4 Q99941	Q99941 homo sapien
32	119	6.2	700	4 Q13289	Q13289 homo sapien
33	118.5	6.1	181	11 Q62281	Q62281 mus musculus
34	118.5	6.1	590	3 Q08490	Q08490 saccharomyc
35	118.5	6.1	1151	5 Q9XUR2	Q9XUR2 caenorhabdi
36	118	6.1	378	10 Q23964	Q23964 helianthus
37	118	6.1	547	3 Q9Y7Y1	Q9Y7Y1 schizosacch
38	117.5	6.1	427	10 Q40587	Q40587 nicotiana t
39	117.5	6.1	601	5 Q76379	Q76379 echinococcu
40	117.5	6.1	656	5 Q9Y114	Q9Y114 drosophila
41	117.5	6.1	1604	4 Q95643	Q95643 homo sapien
42	117.5	6.1	1792	4 Q15129	Q15129 homo sapien
43	117	6.1	355	3 P78962	P78962 schizosacch
44	117	6.1	703	4 Q14343	Q14343 homo sapien
45	117	6.1	1230	5 Q20626	Q20626 caenorhabdi

ALIGNMENTS

RESULT 1

Q16946 ID Q16946 PRELIMINARY; PRU; 378 AA.
AC Q16946;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE APCREB2.
GN APCREB2.
OS Aplysia californica (California sea hare).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidea;
OC Aplysiidae; Aplysia.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-CENTRAL NERVOUS SYSTEM;
RX MEDLINE; 96107336
RA BARTSCH D., GHIRARDI M., SKEHEL P.A., KARL K.A., HERDER S.P., CHEN M.,
RA BAILEY C.H., KANDEL E.R.;
RT "Aplysia CREB2 represses long-term facilitation: relief of repression
RT converts transient facilitation into long-term functional and
RT structural change."
RL Cell 83:979-992(1995).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER BZIP PROTEINS.
DR EMBL; U40851; AAA92437.1; -
DR PROSITE; PS00036; BZIP_BASIC; 1.
DR PFAM; PF00170; BZIP; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 378 AA; 41692 MW; A3595B6E CRC32;

Query Match 99.7%; Score 1923; DB 5; Length 378;
Best Local Similarity 100.0%; Pred. No. 1.1e-123;
Matches 378; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MELDLWSEDFQLAREWGLEMPVQTDGQDLKSTRHGGSLSLQPGATLKLEPFE 60
|||||
Db 1 MELDLWSEDFQLAREWGLEMPVQTDGQDLKSTRHGGSLSLQPGATLKLEPFE 60
|||||
Qy 61 DVLGAENWESSDLGSLFDALGDGNHERLHPFESNLLEFTSLTPDDSTVSKDLSLTLOFP 120
|||||
Db 61 DVLGAENWESSDLGSLFDALGDGNHERLHPFESNLLEFTSLTPDDSTVSKDLSLTLOFP 120
|||||

QY	121	TOPNIPLYASHGAEDFSAETEFENHLSPDPEQVAPVINLEPVELTASHMTVISPDL	180
Db	121	TOPNIPLYASHGAEDFSAETEFENHLSPDPEQVAPVINLEPVELTASHMTVISPDL	180
QY	181	LGMELASESLTTFELDFVNFNDSAVGSIIGGAPELLGSPLSVDDVESTISFSGDPSPETS	240
Db	181	LGMELASESLTTFELDFVNFNDSAVGSIIGGAPELLGSPLSVDDVESTISFSGDPSPETS	240
QY	241	QSSIIESSPELYKVIYSTSIDASKRFSYRSRKSQSVKTSDAKAPRTRTPAQVPVEH	300
Db	241	QSSIIESSPELYKVIYSTSIDASKRFSYRSRKSQSVKTSDAKAPRTRTPAQVPVEH	300
QY	301	VIMEHLDDKRRKLQNKNAAIRYMKKKGEAGGKGEQELBELNKKTKVKDDLQREIK	360
Db	301	VIMEHLDDKRRKLQNKNAAIRYMKKKGEAGGKGEQELBELNKKTKVKDDLQREIK	360
QY	361	YMKNLAMEDVCKAKGIOLK	378
Db	361	YMKNLAMEDVCKAKGIOLK	378

RESULT 2

Q61328	PRELIMINARY;	PRT;	381 AA.
ID	Q61328		
AC	Q61328;		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)		
DT	01-NOV-1999 (TrEMBLrel. 12, Last annotation update)		
DE	ACTIVATING TRANSCRIPTION FACTOR 4 (ACTIVATING TRANSCRIPTION FACTOR 4 (ATF4)).		
DE	ATF4 OR ATF-4.		
GN	Mus musculus (Mouse).		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
OC	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
[1]			
RN	SEQUENCE FROM N.A.		
RC	STRAIN-129/SV;		
RC	MEDLINE; 92066493.		
RT	MELNICKI L.M., PRUITT S.C.;		
RT	"Isolation and nucleotide sequence of a murine CDNA homologous to		
RT	human activating transcription factor 4.";		
RL	Nucleic Acids Res. 19:6332-6332(1991).		
[2]			
RN	SEQUENCE OF 119-381 FROM N.A.		
RP	NISHIZAWA M., NAGATA S.;		
RA	Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.		
[3]			
RN	SEQUENCE OF 119-381 FROM N.A.		
RP	MEDLINE; 92183900.		
RA	NISHIZAWA M., NAGATA S.;		
RT	"cDNA clones encoding leucine-zipper proteins which interact with G-		
RT	CSF gene promoter element 1-binding protein.";		
RL	FEBS Lett. 299:36-38(1992).		
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).		
CC	-1- SIMILARITY: TO OTHER BZIP PROTEINS.		
DR	EMBL; AB012277; CAA43723.1;		
DR	ENBL; X612277; BAA25314.1;		
DR	MGD; MGI:88096; Atf4		
DR	PROSITE; PS00036; BZIP_BASIC; 1.		
DR	PFAM; PF00170; bzip; 1.		
DR	DATA-binding; Nuclear protein.		
SW	SEQUENCE 381 AA; 41810 MW; C8136FE7 CRC32;		

Qy	68	--MESSDGLSDALGDGNHERLHPFFESNLELFTSLTPDDSTVSKDILSSTL-----	111
Db	97	GLASASDTGK--EDAFSGTDWNLKMDLKEFDLALFRMDLFTMPDELUTTLDDTCDLFA	155
Qy	118	-----QFTQPNIPLYASHGAEDFSAETEFENHLSPPDS---DEQVAPVINLEP---	164
Db	156	PLVOETNKEPPQTVN-PI-----GHL--PESLIKVDQVAPFTFLQPPFC	196
Qy	165	-----VEL-----TASHMTVISPOGLLGMELASESUTFTTEL	196
Db	197	SPGVLSSTPETHSFSELGSEVDISEGDRKPDSAAYITLIPP-----CVKEE	242
Qy	197	DFVFNDSAVCSIGCAEELGSLPSLVDVESTISFGSPSPETSSQSIIESSELYKVIS	256
Db	243	DTPSDNDS--GICMSPESYLGSP-----QHSPTFSR-----APPD--NLPS	279
Qy	257	TSSDASAKRFSPYSRSKSKQSVKTSDAKAPKTRTPAQVPVPEHVMEHLDKDRKKLQ	316
Db	280	PGSGRSGPRPKPYD-----PGVSLTAKVKTEKLDKKLKKMEQN	318
Qy	317	KNAAIRYFMKKKGAEQIGKEBEOLEBLNTKTKYVDDDLQRETKYKWNLMEDVCRAKG	374
Db	319	KTAATRYRQKKRAEQEALTGECKELEKKNEALKEDASLAKETQYLKDLIEEVRKRG	376

RESULT

3

Q9W610

PRELIMINARY;

PRT;

354 AA.

AC

Q9W610;

DT

01-NOV-1999 (TrEMBLrel. 12, Created)

DT

01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT

01-NOV-1999 (TrEMBLrel. 12, Last annotation update)

DE

ATF4.

DE

ATF4.

GN

ATF4.

OS

Gallus gallus (Chicken).

OC

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;

OC

Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

RC

RP

SEQUENCE FROM N.A.

RC

TISSUE=EYE.

KA

MURAYA T., SUZUKI H., OGINO H., YASUDA K.;

RT

"Isolation and analysis of embryonic expression pattern of chicken

RT

ATF4."

RL

Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.

CC

-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

CC

-1- SIMILARITY: TO OTHER BZIP PROTEINS.

DR

EMBL; AB013138; BAA76466.1; -.

DR

PROSITE; PS000036; BZIP_BASIC; 1.

DR

DNA-binding; Nuclear protein.

SK

SEQUENCE 354 AA; 38990 MW; 6202DDA0 CRC32;

```

Query Match      12.9%; Score 248; DB 11; Length 381;
Best Local Similarity 27.0%; Pred. No. 1.5e-09;
Matches 113; Conservative 40; Mismatches 125; Indels 140; Gaps

QY 19 EMPVYOT-DQFGDLKSTRHGSDLSLQ-----GATLKLEPFEEDVLG-AEW----- 67
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 37 ELPRQSVGGGLDPPFQSGIGLGAESLGLDDYLEVAKHLKPHGFSDDKAGSSWPAMD 96
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```



```
Db 232 -----QISPTNSLSPNDNP-----FPTDACSRSKPYDHPAEKVASAK 272
QY 281 TSDAKAPKTRTPAQPVPPEHVMEHDKDRKKLQNKNAAIRYMKKKGEAQIGKEEQE 340
Db 273 VKGEK-----KIDKKLKKMEQNKTAATRYQRKKRAEQEALSSECRD 313
QY 341 LEELNTKTKVDDLOREIKYMKNLMDVCCKAK 373
Db 314 LBQKNQALKERADSLSKETQYKLDLIEEVRKAK 346

RESULT 4
QY2D1 PRELIMINARY; PRT; 182 AA.
AC QY2D1;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE LEUCINE-ZIPPER PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA KOHROKI J., TANAKA K.;
RT "cDNA clone encoding leucine-zipper protein.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: TO OTHER BZIP PROTEINS.
DR EMBL; AB021663; BAA78477.1; -
DR PROSITE; PS00036; BZIP_BASIC; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 182 AA; 20428 MW; 61A81591 CRC32;

Query Match 9.8%; Score 188.5; DB 4; Length 182;
Best Local Similarity 30.8%; Pred. No. 6.4e-06;
Matches 45; Conservative 30; Mismatches 56; Indels 15; Gaps 2;

QY 234 PSSPTSSSISSPELYKVIYSTSSIDASKRFSYRSKSKQSVKTSIDAKPRKTRTP 293
Db 39 PSLPLSLFSLPQPPVL-----DTLDLLALYCRNEAQEEVGMPLPPQPPPPP 89
QY 294 AQPVE-----HVIMEHDKDRKKLQNKNAAIRYMKKKGEAQIGKEEQEELNTK 347
Db 90 SPQPQSLAPYPHPATTRGDRKKRQNKSAALRYQRKKRAEGEAGECGLEARNRE 149
QY 348 LKTKVDDLOREIKYMKNLMDVCCKAK 373
Db 150 LKRAESVEREIQYVKDLLIEYKAR 175

RESULT 5
O70191 PRELIMINARY; PRT; 84 AA.
AC O70191;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE ATFX (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RA NISHIZAWA M., NAGATA S.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92183900.
RA NISHIZAWA M., NAGATA S.;
RT "cDNA clones encoding leucine-zipper proteins which interact with g-
```

```
RT CSF gene promoter element 1-binding protein.";
RL FEBS Lett. 299:36-38(1992).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: TO OTHER BZIP PROTEINS.
DR EMBL; AB012276; BAA25313.1; -
DR PROSITE; PS00036; BZIP_BASIC; 1.
DR PFAM; PF00170; BZIP; 1.
KW DNA-binding; Nuclear protein.
FT NON_TER 1
SQ SEQUENCE 84 AA; 9893 MW; 04BE945C CRC32;

Query Match 8.4%; Score 162; DB 11; Length 84;
Best Local Similarity 43.3%; Pred. No. 0.00015;
Matches 29; Conservative 22; Mismatches 16; Indels 0; Gaps 0;

QY 307 DKDKRKKLQNKNAAIRYMKKKGEAQIGKEEQEELNTKTKVDDLOREIKYMKNL 366
Db 11 DRQKKRQNKSAALRYQRKKRAEGEAGECGLEARNRELRRAESVEREIQYVKDLL 70
QY 367 EDVCCKAK 373
Db 71 IEVYKAR 77

RESULT 6
Q23272 PRELIMINARY; PRT; 488 AA.
AC Q23272;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JAN-1999 (TRENBLrel. 09, Last annotation update)
DE ZC376.7 PROTEIN.
GN ZC376.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditida;
OC Rhabditina; Rhabditoidae; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA MCMURRAY A.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z77136; CAB00883.1; -
SQ SEQUENCE 488 AA; 56019 MW; 80F3CB5D CRC32;

Query Match 7.7%; Score 148.5; DB 5; Length 488;
Best Local Similarity 24.1%; Pred. No. 0.012;
Matches 91; Conservative 53; Mismatches 115; Indels 119; Gaps 21;

QY 21 PVVQTGQ--FGDLKSTSRHGDESLSLQPGQATLKLEPFEEVDY-----LGAEN-----M 68
Db 184 PLFQSPKSAIDQLTGTSR---IDEXGMPQDR--KLKSFEMDIEQSKAVDWEAHHYL 238
QY 69 ESSDLGSFLDALGDNHERLHPESNLLFTSLTPDDSTVSKDILSSTLQFTQPVN--- 125
Db 239 ESDD-----DVFKRPEAF---FKPEPMINTSSDS-----LMTSSTSPDSGSLYD 281
```



```
DE ACTIVATING TRANSCRIPTION FACTOR 2
DE (CYCLIC AMP RESPONSE ELEMENT DNA-BINDING PROTEIN ISOFORM 2).
GN ATF2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92123199.
RA GEORGOPOULOS K., MORGAN B.A., MOORE D.D.;
RT "Functionally distinct isoforms of the CRE-BP DNA-binding protein
RT mediate activity of a T-cell-specific enhancer.";
RL Mol. Cell. Biol. 12:747-757(1992).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER BZIP PROTEINS.
DR EMBL; S76659; AAB21129.1; -.
DR HSSP; P03069; 2DGC.
DR MGD; MGI:109349; ATF2.
DR PROSITE; PS00036; BZIP_BASIC; 1.
DR PFAM; PF00170; bZIP; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 358 AA; 38978 MW; D4B31477 CRC32;

Query Match 7.1%; Score 137; DB 11; Length 358;
Best Local Similarity 23.2%; Pred. No. 0.05; Mismatches 45; Indels 60; Gaps 11;
Matches 70; Conservative 45;

Qy 104 DDSTVSKDLSLTQFPQPNIPLY---ASHGAEDFSAETEFENHLSPPDPSPEQVAPVI 160
Db 10 NDSVIVADQTTPTFLKNCVEVGLFELASPFENEFKASEDDIKKPLDLSPLATPII 69
Qy 161 NL---EPVELTASHMTVISPDGLLGMELASESLTFTDELDFVNFNDSAVSIGGAELLG 217
Db 70 RSKTEEPSVWETHQDSPLH-----PESTT-----SDEKLV 101
Qy 218 SPLSVDDVESITSPGSPSPSTOSSIIESPELYKVIS-----TSSIDASKRFPYSR 271
Db 102 RPVTM---VPSVPGIIPGSPSPQVQS---EAKMLKAAALTOQHPVTVNGDTVKHGSGILVR 156
Qy 272 SSKSKQSVKTSDAKAPKTRTPAQP---VPE-----HVIMEHLDKDKRKLQ-NKN 318
Db 157 TQSESRPQSLQQPATSTETPASPATHTPTQNTSGRRRAANEDDPDEKRRKFLERNRA 216
Qy 319 AAIRYRMKKGEAGQIKGEQEELNKLTKTKYDDDLQREIKYMKNLN--EDVCKKAGIQ 376
Db 217 AASRCRKRVVQVQSLKKAEDLSNLQLOQSEVTLRNEVAQLKQLLAHKDCPVTAMQ 276
Qy 377 LK 378
Db 277 KK 278

RESULT 10
Q64090 PRELIMINARY; PRT; 456 AA.
AC Q64090;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE ACTIVATING TRANSCRIPTION FACTOR 2
DE (CYCLIC AMP RESPONSE ELEMENT DNA-BINDING PROTEIN ISOFORM 1).
GN ATF2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92123199.
RA GEORGOPOULOS K., MORGAN B.A., MOORE D.D.;
RT "Functionally distinct isoforms of the CRE-BP DNA-binding protein
RT mediate activity of a T-cell-specific enhancer.";
RL Mol. Cell. Biol. 12:747-757(1992).
```

```
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER BZIP PROTEINS.
DR EMBL; S76657; AAB21128.1; -.
DR HSSP; P03069; 2DGC.
DR MGD; MGI:109349; ATF2.
DR PROSITE; PS00036; BZIP_BASIC; 1.
DR PFAM; PF00170; bZIP; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 456 AA; 48931 MW; 8C79E121 CRC32;

Query Match 7.1%; Score 137; DB 11; Length 456;
Best Local Similarity 21.7%; Pred. No. 0.068; Mismatches 57; Indels 82; Gaps 15;
Matches 85; Conservative 57;

Qy 52 TLKLEPFEEED-VLGAEMMESSDLGSFL---DALGONHERLHFPESNLLLEFTSLITPDDST 107
Db 2 TLKFGPARNDSVIVAD--QTPTPTFLKNCVEVGLFELASPFEN--EFKKAASEDDIKK 56
Qy 108 VSKDILSSTLQFPQPVNIPLYASHGAEDFSAETEFEN-----HLSPPDPSPEQVAPVIN-- 161
Db 57 MPLDL-----SPLATPIIRSKIEEPSVVVETTHQDSPLPHPESTTTSDEKEVPLAQT 107
Qy 162 -----LEPVELTASHMTVISPDGLLGMELASESLTFTDELDFVNFNDSAVSIGGAEL 215
Db 108 QPSTAIVPASLQVNPVLLTSSDSSVITQOAVPSTSTVITQAPSSNNRPVVPFGPPL 167
Qy 216 L-----GSPLSVDD-----VESTISFGSPSPSTSSQSSIIIES 247
Db 168 LHLHPNGQTPMVAIPASITSSNNVHPAAVPLVPVMTVPSPGICPGSPSPQVQS---EA 224
Qy 248 SPELYKVIS-----TSSIDASKRFPYSRSKSKQSVKTSDAKAPKTRTPAQP---VP 298
Db 225 KMLKAAALTOQHPVTVNGDTVKHGSGILVRTQSESRPQSLQQPATSTTETPASPAHTTP 284
Qy 299 E-----HVIMEHLDKDKRKLQ-NKNAIRYRMKKKGAEQAIKGEQEELNKL 348
Db 285 QTQNTSGRRRAANEDDPDEKRRKFLERNRAASRCRKRVVQVQSLKKAEDLSNLQLO 344
Qy 349 KTKVDDLQREIKYMKNLN--EDVCKKAGIQLK 378
Db 345 QSEVTLRNEVAQLKQLLAHKDCPVTAMQK 376

RESULT 11
O93602 PRELIMINARY; PRT; 487 AA.
ID O93602;
AC O93602;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE ACTIVATING TRANSCRIPTION FACTOR 2.
GN ATF2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RA HUGUIER S., BAGUET J., PEREZ S., VAN DAM H., CASTELLAZZI M.;
RT "Transcription factor ATF2 cooperates with vjun to promote growth-
RT factor independent proliferation in vitro and tumor formation in
RT vivo.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER BZIP PROTEINS.
DR EMBL; Y17724; CAA76838.1; -.
DR HSSP; P08047; ISP2.
DR PROSITE; PS00036; BZIP_BASIC; 1.
DR PFAM; PF00170; bZIP; 1.
DR PFAM; PF00096; zf-C2H2; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 487 AA; 52406 MW; 52Fb7330 CRC32;
```

```

Query Match          7.1%; Score 137; DB 13; Length 487;
Best Local Similarity 21.3%; Pred. No. 0.074;
Matches 84; Conservative 60; Mismatches 163; Indels 88; Gaps 15;

QY 52 TLKLEPFEED-VLGAEMWESSDLGSFL---DALGDNHERLHPFESLNLEFTSLITPDOST 107
DB 34 TLKFGPARNDSVIAD--QTPTPTFRFLKNCVEEVLGNELASPFEN---EFKASEDDIKK 88

QY 108 VSKDILSTLQFPQPVPNIPLYASHGAEDFSAETEFENHLSLSP-PDSPEQVAPVINLE--- 163
DB 89 MPLDL-----SPLATPIIRNKIEEPSVVEETHQDSPLPHPESTTNDEKVESLQOTA 139

QY 164 -----PVELTASHMTVISPDLGLGMELASESLFTTDLDFVNFNDSVAGSIGGAEL 215
DB 140 QPTSAIVRPASLPVNPVLLTSSDSSVLIQQAIVPSTSTVITQAPSSNRRIIVPVGPFPL 199

QY 140 QPTSTIVRPASLPVNPVLLTSSDSSVLIQQAIVPSTSTVITQAPSSNRRIIVPVGPFPL 199
DB 200 LLHLPNGQTMVPAIPASITSSNVHVPAAVPLVRPVTWVPSPGIPGPSSPQVQSEAKLR 259

QY 216 L-----GSPLSVDD-----VESTISFGSPSPSPETSQS----- 242
DB 200 LLHLPNGQTMVPAIPASITSSNVHVPAAVPLVRPVTWVPSPGIPGPSSPQVQSEAKLR 259

QY 243 ---SIIESSPELYKVISSIDASK-----RFSYRSKSKSQSVKTSDAKAPKTRT 292
DB 260 LKAALTOQHFOV-----TNGDTAKGHPGLVTRTQSEPRPSLQQAATSTETPASPQA 313

QY 293 PAQVPV-----EHVIMEHLDDKDRKKLQ-KNNAAIRYRMKKKGAEQIGKEEELFN 345
DB 314 PTQQTPTNGRRRAANEDPDEKRRKFLERNRAAASRCRQKRVVQSLKKAEDLSLN 373

QY 346 TLKTKVDDQLQREIKYKMKML--EDVCKAKGIQK 378
DB 374 GOLQNEVTLRNEVAQLKQLLAHKDCPVTAMQKK 408

RESULT 12
Q91576 PRELIMINARY; PRT; 486 AA.
AC Q91576;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE ACTIVATING TRANSCRIPTION FACTOR 2.
GN ATF2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae;
OC Xenopus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95180723.
RA VILLARREAL X.C.; RICHTER J.D.;
RT "Analysis of ATF2 gene expression during early Xenopus laevis
development."
RL Gene 153:225-229(1995).
DR EMBL; U16158; AAA69518.1; -.
DR HSP; P08047; 1SP2.
DR PFAM; PF00170; bZIP; 1.
DR PFAM; PF00096; zf-C2H2; 1.
SQ SEQUENCE 486 AA; 52378 MW; 4F7C23E3 CRC32;

Query Match          7.0%; Score 135; DB 13; Length 486;
Best Local Similarity 21.5%; Pred. No. 0.1;
Matches 88; Conservative 56; Mismatches 148; Indels 118; Gaps 17;

QY 52 TLKLEPFEED-VLGAEMWESSDLGSFL---DALGDNHERLHPFESLNLEFTSLITPDOST 107
DB 34 TLKFGPARNDSVIAD--QTPTPTFRFLKNCVEEVLGNELASPFEN---EFKASEDDIKK 87

QY 108 VSKDILSTLQFPQPVPNIPLYASHGAEDFSAETEFENHLSLSPDPSPQV-----APVI 160
DB 88 MPLDL-----SPLATPIIRNKIEEPSVVEETHQDSPLPHPESTTSDQHNQVPLA 136

```

```

QY 161 N-----LEPVELTASHMTVISPDLGLGMELASESLFTTDLDFVNFNDSVAGSIGGA 212
DB 137 QTAOPTSAIVRPASLPVNPVLLTSSDSSVLIQQAIVPSTSTVITQAPSTNSSIVTLPGP 196

QY 213 EELL-----GSPLSVDD-----VESTISFGSPSPSPETSQSSI 244
DB 197 FVLLHLPLSGQTMVPAIPASITSSNVHVPAAVPLLRPLTWMVSPGIPGPSSPQVQS-- 254

QY 245 IESSPELYKVISSIDASKRFS-----PYSRSKSKSQSVKT 281
DB 255 -----EAKMRFNAATQQHPPVTNGDTVKGHSGFAITQSEESRPSLO 298

QY 282 SDAKAPKTR-TPAQVPE-----HVIMEHLDDKDRKKLQ-KNNAAIRYRMKKKG 330
DB 299 QPATSTTETRPSPAQPTQTQSTSGRRRAANEDPDEKRSKLIQNRRAAASRCRQKRVK 358

QY 331 AQGKIGEOELELNTKLTKYKVDLQREIKYKMKML--EDVCKAKGIQK 378
DB 359 VQSLKKADELISLNGYLQNEVTLRNEVAQLKQLLAHKDCPVTAMQKK 408

RESULT 13
Q62870 PRELIMINARY; PRT; 487 AA.
AC Q62870;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE CAMP RESPONSE ELEMENT BINDING PROTEIN 1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=SPRAGUE DAWLEY; TISSUE=BRAIN;
RA MURAMATSU S.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
CC -|- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -|- SIMILARITY: TO OTHER BZIP PROTEINS.
DR EMBL; U38938; AAA93263.1; -.
DR HSP; P08047; 1SP2.
DR PROSITE; PS00036; BZIP_BASIC; 1.
DR PFAM; PF00170; bZIP; 1.
DR PFAM; PF00096; zf-C2H2; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 487 AA; 52286 MW; 29CEDF55 CRC32;

Query Match          7.0%; Score 134; DB 11; Length 487;
Best Local Similarity 21.9%; Pred. No. 0.12;
Matches 86; Conservative 58; Mismatches 166; Indels 82; Gaps 16;

QY 52 TLKLEPFEED-VLGAEMWESSDLGSFL---DALGDNHERLHPFESLNLEFTSLITPDOST 107
DB 34 TLKFGPARNDSVIAD--QTPTPTFRFLKNCVEEVLGNELASPFEN---EFKASEDDIKK 88

QY 108 VSKDILSTLQFPQPVPNIPLYASHGAEDFSAETEFENHLSLSP-PDS---PEQVAPVIN-- 161
DB 89 MPLDL-----SPLATPIIRNKIEEPSVVEETHQDSPLPHPESTTNDEKEIPLAOTA 139

QY 162 -----LEPVELTASHMTVISPDLGLGMELASESLFTTDLDFVNFNDSVAGSIGGAEL 215
DB 140 QPTSAIVRPASLPVNPVLLTSSDSSVLIQQAIVPSTSTVITQAPSSNRRIIVPVGPFPL 199

QY 216 L-----GSPLSVDD-----VESTISFGSPSPSPETSQSSI 247
DB 200 LLHLPNGQTMVPAIPASITSSNVHVPAAVPLVRPVTWVPSPGIPGPSSPQVQS---EA 256

QY 248 SPELYKVIS-----TSSIDAKRFSYRSKSKSQSVKTSDAKAPKTRTPAQP---VP 298
DB 257 KMLKAALTOQHPPVTNGDTVKGHSGVLRQASRPSLQQAATSTTETPASPAHTTP 316

QY 299 E-----HVIMEHLDDKDRKKLQ-KNNAAIRYRMKKKGAEQIGKEEELNKL 348

```

```
Db 317 QTONTSGRRRAADDPDEKRRKFLERNRAASRCRQRKVVQVQSLERKAEDLSSINGQL 376
Qy 349 KTKVDLQREIKYKNLM--EDVCKAKGIQLK 378
Dt 377 QSEVTLRNEVAQLKQLLAHKDCPVTAMQK 408

RESULT 14
Q64089 PRELIMINARY; PRT; 448 AA.
AC Q64089;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE ACTIVATING TRANSCRIPTION FACTOR 2
DE (CYCLIC AMP RESPONSE ELEMENT DNA-BINDING PROTEIN ISOFORM 3).
GN ATF2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92123199.
RA GEORGOPOULOS K., MORGAN B.A., MOORE D.D.;
RT "Functionally distinct isoforms of the CRE-BP DNA-binding protein
RT mediate activity of a T-cell-specific enhancer.";
RL Mol. Cell. Biol. 12:747-757(1992).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER BZIP PROTEINS.
DR EMBL: S76655; AAB21127.1; -.
DR HSSP: P03069; 2DGC.
DR MGD: MGI:109349; Atf2.
DR PROSITE: PS00036; BZIP_BASIC; 1.
DR PFAM: PF00170; bzip; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE .448 AA; 48267 MW; 6698E12B CRC32;
```

```
Query Match 6.9%; Score 133; DB 11; Length 448;
Best Local Similarity 21.0%; Pred. No. 0.12;
Matches 76; Conservative 53; Mismatches 157; Indels 76; Gaps 12;

Qy 78 DALGDNHERLHPFESNLEFTSLITPDSTVSKDILSTLQFTQPVNPIYASHGAEDF 137
Db 22 EEVGLFENELASPFEN--EFKASEDDIKMPLDL-----SPLATPIIRSKIEEPS 69
Qy 138 SAETEFEN-----HLSPPDSPQVAPVIN-----LEPVELTASHMTVISPDLGLGME 185
Db 70 VVETHQDSPLPHPSESTTSDEKEVPLAQTAQPSAIVRPASLQVPNVLLTSSDSVITQQ 129
Qy 186 LASESLTFTELDFVNFNDSAVGSGGABELL-----GSPLSVDD----- 224
Db 130 AVPSTSTVITQAPSSNRPIVPVGPFPPLLHLPGNGTMPVAIPASITSSNVHVPAAPV 189
Qy 225 -----VESTISFGSPSPSTSSSIIESPELYKVIS-----TSSIDASKRFSYSR 271
Db 190 LVRPVTWVPSPVPGIPGSPSPQVQS--EAKMKLKAALTQHPVNGDTVKHGSGGLVR 246
Qy 272 SSKSKQSVKTSDAKAPKTRTPAQP--VPE-----HVIMEHDKDKRKKLQ--NKN 318
Db 247 TQSESRPQSLQQQATSTSTETPASPAPHTPQTQNTSGRRRAANEDDPDEKRRKFLERNR 306
Qy 319 AAIRYRMKKKGAEQIKGEQEELELNTKLTKYDDQLQREIKYKNLM--EDVCKAKGIQ 376
Db 307 AASRCRQRKVVQVQSLERKAEDLSSINGQLQSEVTLRNEVAQLKQLLAHKDCPVTAMQ 366
Qy 377 LK 378
Db 367 KK 368
```

RESULT 15

```
O08592 PRELIMINARY; PRT; 2390 AA.
AC O08592;
DT 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE VERSICAN V0 ISOFORM (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR KYOTO;
RA LEMIRE J.N., BRAUN K.R., MAUREL P., MARGOLIS R.U., SCHWARTZ S.M.,
RA WIGHT T.N.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR KYOTO;
RA MILEV P., MAUREL P., CHIBA A., MEVISSSEN M., POPP S., YAMAGUCHI Y.,
RA MARGOLIS R.K., MARGOLIS R.U.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 2073-2115 FROM N.A.
RC TISSUE=NORMAL KIDNEY;
RA PYKE C., KRISTENSEN P., OSTERGAARD P.B., OTURAI P.S., ROMER J.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF062402; AAC40166.1; -.
DR EMBL: U75306; AAB51125.1; -.
DR HSSP: P01132; IEPH.
DR PROSITE: PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE: PS01187; EGF_CA; 1.
DR PFAM: PF00008; EGF; 2.
DR PFAM: PF00059; Lectin_c; 1.
DR PFAM: PF00084; sushi; 1.
KW Glycoprotein; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 2390 AA; 261150 MW; DBE8E4A4 CRC32;
```

```
Query Match 6.8%; Score 131; DB 11; Length 2390;
Best Local Similarity 21.4%; Pred. No. 1.5;
Matches 82; Conservative 53; Mismatches 147; Indels 102; Gaps 15;

Qy 20 MPVQTDQGFGLKTSRRHGDESLQPO--CATLKLEPFEDVLGAENWESSDLGSLFD 78
Db 1597 IPGIQTDLD--SDIPLGSHGSEESLEVOEKYEATNLSPTTEAFDGS-----D 1644
Qy 79 ALGDNHERLHPFESNLEFTSLITPDSTVSKDILSTLQFTQPVNPIYASHGAEDFS 138
Db 1645 ALPAGHTQAIYNES-----VTPSDGKQPEDI---SFSFAT---GIPV-----SS 1682
Qy 139 AETEFENHLSPPDSPQVAPVINLEPVELTASHMTVISPDLGLGMEASESLTTELDF 198
Db 1683 TETEL--NTFFPVTSLHIPSKLTTTASPEIDKPNIBALSDDIF-----ESSTLSGQA 1734
Qy 199 VNFNDSAVGSGGABELLGSPLSDVDVESTISFGSPS----- 235
Db 1735 IADQSEVSTLGHLEK-----TQEEYEK--KYGFSQPEFFSGVGVEFTDAPYVSG 1787
Qy 236 -----SPETSQSSIIESPELYKVISTSSIDASKRFSYSKSKSQSVKTSDAKAPRK 289
Db 1788 RTYSVAQPLTEFPNVVQSDSTHYTEATSAVSSVTLSPTSPSPSPVYIDSGVSEFTEV 1847
Qy 290 TETPAQVPEHVIMEHDKDKRKKLQNKNAAIRYRMKKKG-----AQGIK 335
Db 1848 PHKSAQPAPTAASSQKLIEGSEFKYR---ANIEATIKSLGENDHGTSPSPSPALDIS 1904
Qy 336 GEE-----QEELELNTKLTKVD 353
Db 1905 EDDSKPKLLEDLETSTPTTETSQD 1928
```

Search completed: August 8, 2000, 03:26:40
Job time: 6191 sec

```
|||||.....: :||| ||| |||||
705 TCTCCAGAGTTTAAACAAACCCGACCGAGTGGCCCCCTTCACCTCTTAC 656
168 rAlaSerHisMetThrValIleSerProAspGlyLeuLeuGlyMetG 185
555 AACCTCTTCCCCC.....TTTCCCGAGGGGTCTGTCTCCCTCCTC 615
185 luLeuAlaSerGluSerLeuThrPheThrGluLeuLeuAspPheValAsnPhe 201
614 AGATCATTTCTTGTAGTTAG.....: :||| ||| |||||
202 AsnAspSerAlaValGlySerIleGlyAlaGluLeuLeuGlyse 218
591 .....: :||| ||| |||||
218 rProLeuSerValAspValGluSerThrIleSerPheSerGly..... 233
583 TGAAGTGGATATCACTGAAGAGATAGAGAACCCAGACTACACTGCTTACG 534
234 .....ProSerSerProGluThrSer 240
533 TTGCCATGATCCCTCAGTCGATCAAGAGGAGGAGACACCCCTTCAGATAAT 484
241 GlnSerSerIleLeuGluSerProGluLeuThrValIleSerTh 257
483 GATAGTGGCATCTGT...ATGAGCCAGAGTCTCTAT..... 451
257 rSerSerIleAspAlaSerLysArgPheSerProGluThrSer 274
450 .....NTGGGGTCTCTCCAGCACAGCCCTCTACCGAGGGCTCTC 411
274 ySerLysGlnSerValLysThrSerAspAlaLysAlaProArgLysThr 290
410 CAATAGS...AGCTCCCATCTCCAGGTGTCTCTGTGGTCTGCCCGT 364
291 ArgThrProAlaGlnProValProGluHisValIleMet..... 303
363 CCCAAACCTTACGATCTCTCTGGAGAGAGATGGTAGCAGCAAAAGTAAA 314
304 ....GluHisLeuAspLysLysAspArgLysLysLeuGlnAsnLysAsnA 319
313 GGGTGAGAACTGGATAGAGAGCTGAAAGAAATGGAGCAAAACAGACAG 264
319 laAlaIleArgTyrArgMetLysLysGlyGluAlaGlnGlnIleLys 335
263 CAGCCACTAGTACCGCAGAGAGAGGGCGGAGCAGAGGCTTACT 214
336 GlyGluGlnGluLeuLeuGluLeuAsnThrLysLeuLysThrLysVa 352
213 GGTGAGTGCAAGAGCTGGAAGAAAGAACGAGGCTCTTAAAGAGAGGGC 164
352 lAspAspLeuGlnArgGluIleLysTyrMetLysAsnLeuMetGluAspV 369
163 GGATTCCTGGCCAGGAGATCCAGTACCTGAAAGATTGTGATAGAAGAG 114
369 aCysLysAlaLysGly 374
113 TCCCAAGGCAAGGGGG 97
```

seq_name: gb_est19:AI339383

seq_documentation_block: 776 bp mRNA EST 29-DEC-1998
LOCUS AI339383
DEFINITION qt02a04.x1 NCI_CGAP_c4 Homo sapiens cDNA clone IMAGE:1946382 3'
similar to gb:D90209 DNA-BINDING PROTEIN TAXREB67 (HUMAN);, mRNA
sequence.

ACCESSION AI339383
VERSION AI339383.1 GI:4076310
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 776)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Jun 15, 1998 this sequence version replaced gi:3222450.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 464.

FEATURES
source

1..776
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1946382"
/clone_lib="NCI_CGAP_G4"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from 3 pooled
germ cell tumors, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."
BASE COUNT 160 a 203 c 187 g 223 t 3 others
ORIGIN

alignment_scores:
Quality: 216.50 Length: 249
Ratio: 1.580 Gaps: 9
Percent Similarity: 55.020 Percent Identity: 29.317
alignment_block:
US-08-656-811a-1 x AI339383/rev ..

Align seg 1/1 to reverse of: AI339383 from: 1 to: 776
144 GluAsnHisLeuSerProAsp..... 151
750 GAGACTATACAGCCCCCCCCAGCGGTGACACAAATTGGCATCTCCCA 701
152SerProGluGlnValAlaProValIleAsnLeuGluP 164
700 GAAAGTTTACAAAAACCGAGCGGTTGNCNCTTACCATCTACAACC 651
164 roValGluLeuThrAlaSerHisMetThrValIleSerProAspGlyLeu 180
650 CTCTTCACCTTTCCCGAGGGGTCTGTCTCCCTACACAGATCATTCCTTT 601
181*LeuGlyMetGluLeuAlaSerGluSerLeuThrPheThrGluLeuAs 197
600AGTTTAGAGCTGGGAGCTGAA...GTGGATATCACTGAAGGAGA 560
197 pPheValAsnPheAsnAspSerAlaValGlySerIleGlyAlaGluG 214
559 TAGNAAGCCAGACTACACTGCTTACGTGCCATGATC..... 523
214 luLeuLeuGlySerProLeuSerValAspValGluSerThrIleSer 230
522CCTCAGTCATAAAGGAGGAGAC..... 499

231 PheSerGlyProSerSerProGluThrSerGlnSerSerIleIleGluSe 247
 498ACCCCTCAGATATGATAGTGCATCTGT...AT 467
 247 rSerProGluLeuTyrLysValIleSerThrSerSerIleAspAlaSerL 264
 466 GAGCCAGAGTCCATCTG.....GGGTCTC 441
 264 ysArgPheSerProTyrSerArgSerSerLysSerLysGlnSerValLys 280
 440 CTCAGCACAGCCCTCTACCAAGGGCTCTCCAAATAGG...AGCCTCCCA 394
 281 ThrSerAspAlaLysAlaProArgLysThrArgThrProAlaGlnProVa 297
 393 TCTCCAGTGTCTCTGTGGGTCTGCCCGTCCCAACCTTACGATCTCC 344
 297 lProGluHisValIleMet.....GluHisLeuAspLysL 309
 343 TGGAGAGAAGATGGTAGCAGCAAAAGTAAGGCTCAGAACTGGATAAGA 294
 309 ysAspArgLysLysLeuGlnAsnLysAsnAlaIleArgTyrArgMet 325
 293 AGCTGAAAAAATGAGCAAAACAGACAGCAGCCACTAGGTACGCCAG 244
 326 LysLysLysGlyGluAlaGlnGlnIleLysGlyGluGlnGlnGluLeuG 342
 243 AAGAAGGGGGGAGCAGCAGGCTCTTACTGTGTGTCGCAAGAGCTGGA 194
 342 uGluLeuAsnThrLysLeuLysThrLysValAspAspLeuGlnArgGluI 359
 193 AAAGAGNACGAGGCTCTAAAGAGAGAGGGCGGNTCCCTGCCAAGGAGA 144
 359 leLysTyrMetLysAsnLeuMetGluAspValCysLysAlaLysGly 374
 143 TCCAGTACCTCAAGATTGTAGAGAGGTCGCCAGGCAAGGGGG 97

seq_name: gb_est16:A1114565

seq_documentation_block: 11-NOV-1999
 LOCUS A1114565 652 bp mRNA EST
 DEFINITION HAI149 Human fetal liver cdna library Homo sapiens cdna, mRNA
 sequence.
 ACCESSION A1114565
 VERSION A1114565.1 GI:6359910
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 652)
 Yu, Y., Zhang, C., Luo, L., Ouyang, S., Li, W., Wu, J., Zhou, S., Liu, M.
 and He, P.
 Expression profile analysis of a human fetal liver cdna library
 Unpublished (1998)
 Contact: Yongtao Yu
 Department of Hematology
 Beijing Institute of Radiation Medicine
 27 Taiping Road, Beijing 100850, P.R.China
 Tel: 0086-10-68159479
 Fax: 0086-10-68214653
 Email: yyt48@yahoo.com.
 Location/Qualifiers
 1. .652
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Human fetal liver cdna library"
 /tissue_type="liver"
 /dev_stage="fetal"
 /lab_host="MC1061/P3"
 /note="Vector: pCDNA1"

FEATURES

source
 204 a 137 c 177 g 128 t 6 others
 BASE COUNT
 ORIGIN

alignment_scores:
 Quality: 213.00 Length: 224
 Ratio: 1.805 Gaps: 7
 Percent Similarity: 52.679 Percent Identity: 31.696
 alignment_block:
 US-08-656-811a-1 x A1114565 ..
 Align seg 1/1 to: A1114565 from: 1 to: 652
 158 ProValIleAsnLeuGluPro.....ValGluLeu 167
 13 CCAGTACTGCAGACTATGTCATGTTCAAGTGGGACCTCAAGCTGCACAAG 62
 167 uThrAlaSerHisMetThrValIleSerProAspGlyLeuLeuGlyGly 184
 63 TACCCGCTTAACATATG.....CCTACTTCCGTGGTGGGATGC 100
 184 etGluLeuAlaSerGluSerLeuThrPheThrGluLeuAspPheValAsn 200
 101 TGGAGATGCCTTTGATCCCTCAGTCATTAAGGAGGAGACACACCCCTTCA 150
 201 PheAsnAspSerAlaValGlySerIleGlyGlyAlaGluGluLeuLeuG 217
 151 GATAATGATAGT.....GGCATCTGTATGAGCCAGAGTCTCTATCTGGG 194
 217 ySerProLeuSerValAspValGluSerThrIleSerPheSerGlyP 234
 195 GTCCTCT.....C 202
 234 toSerSerProGluThrSerGlnSerSerIleIleGluSerSerProGlu 250
 203 AGCACAGCCCTTACC...AGGGCTCTCCAATAGAGCCTCCCA... 246
 251 LeuTyrLysValIleSerThrSerSerIleAspAlaSerLysArgPheSe 267
 247TCTCCAGGTGTTCTGTGTGGTCTGCCCTGCCCAA 281
 267 rProTyrSerArgSerSerLysSerLysGlnSerValLysThrSerAspA 284
 282 ACCTTACGATCTCTCTGGAGAGAGATGGTAGCAGCAAAAGATAAGGGT. 330
 284 laLysAlaProArgLysThrArgThrProAlaGlnProValProGluHis 300
 330 330
 301 ValIleMetGluHisLeuAspLysLysAspArgLysLysLeuGlnAsnLy 317
 331GAGAACTGCATAGAAGCTGAAAAAATAATGGAGCAAAACAA 371
 317 sAsnAlaAlaIleArgTyrArgMetLysLysLysGlyGlyAlaGlnGlyI 334
 372 GACAGCAGCCACTAGTACCCGCAAGAAGAGGGCGGAGAGAGGCTC 421
 334 leLysGlyGluGlnGlnLeuGluGluLeuAsnThrLysLeuLysThr 350
 422 TTACTGGTGGTGCAGAGAGCTGGAAAGAGACAGCAGGCTCTTAAAGAG 471
 351 LysValAspAspLeuGlnArgGluIleLysTyrMetLysAsnLeuMetG 367
 472 AGGGCGGATTCCTGCGCAAGGAGATCCAGTACCTGAAAGATTGTATAGA 521
 367 uAspValCysLysAlaLysGly 374
 522 AGAAGTCCGCAAGGCAAGGGGG 543
 seq_name: gb_est22:A1598953

seq_documentation_block:
 LOCUS A1598953 577 bp mRNA EST 21-APR-1999
 DEFINITION EST250656 Normalized rat embryo, Bento Soares Rattus sp. cdna clone
 REMD233 3' end, mRNA sequence.

```

ACCESSION   AI598953
VERSION     AI598953.1  GI:4608001
KEYWORDS    EST.
SOURCE      Rattus sp.
ORGANISM    Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 577)
AUTHORS    Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
TITLE      Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
JOURNAL    Unpublished (1998)
COMMENT    On Jan 19, 1998 this sequence version replaced gi:2284562.
            Other_ESTs: TC74232
            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
            Location/Qualifiers
            source          1..577
                           /organism="Rattus sp."
                           /db_xref="taxon:10118"
                           /clone="RBMZ33"
                           /clone_lib="Normalized rat embryo, Bento Soares"
                           /dev_stage="embryo 8, 12, 18 dpc"
                           /note="Vector: pT73Pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT   132 a 151 c 124 g 170 t
ORIGIN
alignment_scores:
  Quality: 211.00      Length: 156
  Ratio: 2.153         Gaps: 4
  Percent Similarity: 62.821  Percent Identity: 37.821
alignment_block:
  US-08-656-811A-1 x AI598953/rev ..
  Align seg 1/1 to reverse of: AI598953 from: 1 to: 577
233 GlyProSerProGluThrSerGlnSerSerIlelleGluSerSerPr 249
||||| :|||||: :|||:
560 GGAGATAGGAAGCCTGACTCT.....GCTGCTTATATTACTTAACCCC 517
||||| :|||||: :|||:
249 oGluLeuTyrlsValIleSerThr...SerSerIleAspAlaSerLysA 265
||||| :|||||: :|||:
516 TCAGTGTGAAGGAGGAGGACCTCCCTCTGATAGTGACAGTGGCATCT 467
||||| :|||||: :|||:
265 rgPheSerProTySerArgSerLysSerLysGlnSerVallysthr 281
||||| :|||||: :|||:
466 GTATGAGCCCTGAGTCCCTACCTGGCTCTCCCAACACAGCCCTCCACC 417
||||| :|||||: :|||:
282 SerAspAlaLys.....AlaProArgLysTh 290
||||| :|||||: :|||:
416 TCCAGGCCCCACCAGACAGCTCCCTCTCCAGGTGTCTCGTGGTTC 367
||||| :|||||: :|||:
290 rArg.....ThrProAlaGlnProValProGluHisValI 302
||||| :|||||: :|||:
366 TCGACCCCAACCTTATGCCACCTGGAGTAGTGACACCTAAAGTGA 317
||||| :|||||: :|||:
302 leMetGluHisLeuAspLysLysAspArgLysLysLeuGlnAsnLysAsn 318
||||| :|||||: :|||:
316 AGACTGAAAGCTTGATAAGAGCTGAAAGATGGACGACAAACAAGACA 267
||||| :|||||: :|||:
319 AlaAlaIleArgTyArgMetLysLysLysGlyGluAlaGlnGlyIleLy 335
||||| :|||||: :|||:
266 GCACCTACTAGTACCGCCAGAAAGAGGCTGACGAGGAGCCCTCAC 217

```

```

335 sgLyGluGluGlnGluLeuGluGluLeuAsnThrLysLeuLysThrLysV 352
||||| :|||||: :|||:
216 TGGCGAGTGTAAAGAGCTAGAAAGAAAGACGAGGCTCTGAAGAAGAGG 167
||||| :|||||: :|||:
352 alAspAspLeuGlnArgGluLleLysTyMetLysAsnLeuMetGluAsp 368
||||| :|||||: :|||:
166 CAGATTCTCTCGCCCAAGAGATTCTAGTATCTAAAGACCTGTATAGAAG 117
||||| :|||||: :|||:
369 ValCysLysAlaLysGly 374
||||| :|||||: :|||:
116 GTCCGTAAAGCAAGGGG 99
||||| :|||||: :|||:
seq_name: gb_est16:AI101524
seq_documentation_block:
  LOCUS      AI101524      582 bp      mRNA      EST      31-JAN-1999
  DEFINITION EST210813 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
  RBRBR68 3' end, mRNA sequence.
  ACCESSION  AI101524
  VERSION    AI101524
  KEYWORDS   EST.
  SOURCE     Rattus sp.
  ORGANISM   Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 582)
AUTHORS    Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
TITLE      Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
JOURNAL    Unpublished (1998)
COMMENT    Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
            Location/Qualifiers
            source          1..582
                           /organism="Rattus sp."
                           /db_xref="ATCC (inhost):2024131"
                           /db_xref="taxon:10118"
                           /clone="RBRBR68"
                           /clone_lib="Normalized rat brain, Bento Soares"
                           /note="Organ: brain; Vector: pT73Pac; Site_1: EcoRI;
                           Site_2: NotI"
BASE COUNT   133 a 153 c 124 g 172 t
ORIGIN
alignment_scores:
  Quality: 211.00      Length: 156
  Ratio: 2.153         Gaps: 4
  Percent Similarity: 62.821  Percent Identity: 37.821
alignment_block:
  US-08-656-811A-1 x AI101524/rev ..
  Align seg 1/1 to reverse of: AI101524 from: 1 to: 582
233 GlyProSerProGluThrSerGlnSerSerIlelleGluSerSerPr 249
||||| :|||||: :|||:
560 GGAGATAGGAAGCCTGACTCT.....GCTGCTTATATTACTTAACCCC 517
||||| :|||||: :|||:
249 oGluLeuTyrlsValIleSerThr...SerSerIleAspAlaSerLysA 265
||||| :|||||: :|||:
516 TCAGTGTGAAGGAGGAGGACCTCCCTCTGATAGTGACAGTGGCATCT 467
||||| :|||||: :|||:
265 rgPheSerProTySerArgSerLysSerLysGlnSerVallysthr 281

```

```

466 GTATGAGCCCTGAGTCTCTACCTGGGCTCTCCCAACACAGCCCTTCCACC 417
282 SerAspAlaLys.....AlaProArgLysTh 290
416 TCCAGGGCCCCACACAGACAGTCTGCCTTCTCCAGGTGTCTCGGTTC 367
290 rarg.....ThrProAlaGlnProValProGluHisValI 302
366 TCGACCCCAACCTTATGACCACCTGGAGTGTAGTGACAGCTAAAGTGA 317
302 LeMetGluHisLeuAspLysLysAspArgLysLysLeuGlnAsnLysAsn 318
316 AGACTGAAAGCTTGGATGAAGAGCTGAAAGAGATGGACCAAAACNAGACA 267
319 AlaAlaIleArgTyrArgMetLysLysGlyGluAlaGlnGlyIleLys 335
266 GCAGCTACTAGGTACCGCCCAAGAGAGGGCTGAGCAGGAAGCCCTCAC 217
335 sGlyGluGlnGluGlnLeuGluGluLeuAsnThrLysLysLeuLysThLysV 352
216 TGGCGAGTGTAAAGAGCTAGAAAAGAGAACGAGGCTCTGAAAGAGAAGG 167
352 aAspAspLeuGlnArgGluIleLysTyrMetLysAsnLeuMetGluAsp 368
166 CAGATTCTCTGCCAAGAGATTCAGTATCTAAAAGACCTGATAGAAGAG 117
369 ValCysLysAlaLysGly 374
116 GTCCGTAAAGCAAGGGG 99
```

seq_name: gb_est28:AJ396301

seq_documentation_block: 749 bp mRNA EST 01-MAR-2000
LOCUS AJ396301 dkfz426 Gallus gallus cdna clone 25p3r1, mRNA sequence.
DEFINITION AJ396301
ACCESSION AJ396301
VERSION AJ396301.1 GI:7128099

KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus

REFERENCE 1 (bases 1 to 749)
AUTHORS Buerstede,J.M.
TITLE A large database of chicken bursal ESTs as a resource of the analysis of vertebrate gene function
JOURNAL Unpublished (2000)
COMMENT Contact: Buerstede JM
Cellular Immunology
Heinrich-Pette-Institute
Martinistr. 52, 20251 Hamburg, Germany
Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.

FEATURES
Location/Qualifiers
1..749

BASE COUNT 236 a 160 c 181 g 172 t
ORIGIN
/organism="Gallus gallus"
/strain="CB"
/db_xref="taxon:9031"
/clone="25p3r1"
/clone_lib="dkfz426"
/tissue_type="Bursa of Fabricius"
236 a 160 c 181 g 172 t

alignment_scores:
Quality: 210.00 Length: 244
Ratio: 1.603 Gaps: 8
Percent Similarity: 53.689 Percent Identity: 28.279

alignment_block:
US-08-656-811a-1 x AJ396301

Align seg 1/1 to: AJ396301 from: 1 to: 749

```

150 ProAspSerPro.....GluGlnValAlaProValIleAsnLeuG1 163
23 CCTGAATCTCCAATTGGAGCAGATCCCATGGCCCTTTATCTTCCCTCTG 72
163 uproValGluLeuThrAlaSerHisMetThrValIleSerProAspGlyL 180
73 GTCTTTTCCCTCTCCCGAGGTCTCTGACTTCTGT...CCAGACCATT 119
180 euLeuGlyGlyMetGluLeuAlaSerGlu..... 189
120 CATTT...AGTTTGAACCTAGGTAGTGAAGTGGATGTTCTGGAAGGTGA 166
190 .....SerLeuThrPheThrGluLe 196
167 AGGAAACAGGAGGGCCCACTTTCTGGTACTGATCACCAAGTCAGAGAA 216
196 uasppPheValAsnPheAsnAspSerAlaValGlySerIleGlyGlyAlaG 213
217 AGAGGAGGAGAACCATTCGATGATAGT...GGAATATGCATGAGCCAG 263
213 luGluLeuLeuGlySerProLeuSerValAspValGluSerThrIle 229
264 ACTCTTATTTTGGGAACCC..... 283
230 SerPheSerGlyProSerSerProGluThrSerGlnSerSerIleIleG1 246
284 .....CAACATAGTCTCCTACCAATTCTGGATCACCACCAATCA 321
246 uSerSerProGluLeuTyrLysValIleSerThrSerSerIleAspAlaS 263
322 CAACCCG.....TTCCCTACAGATGCCGCTGTGGCT 353
263 erLysArgPheSerProTyrSerArgSerSerLysSerLysGlnSerVal 279
354 CTGTGCGGTCCAAACCATATGACCACCTGCTGAGAAAGGTACGCTCAGCA 403
280 LysThrSerSerAlaLysAlaProArgLysThrArgThrProAlaGlnPr 296
404 AAGGTGAAAGGAGAGAAAG..... 421
296 oValProGluHisValIleMetGluHisLeuAspLysLysAspArgLysL 313
422 .....AAAATAGATAAAAACTGAAAAAGA 446
313 ysLeuGlnAsnLysAsnAlaAlaIleArgTyrArgMetLysLysLysGly 329
447 TGGAGCAAAATAAGACAGCTGCCACACGTTACCGCGCAAAAGAGGGCA 496
330 GluAlaGlnGlyIleLysGlyGluGlnGluLeuGluGluLeuAsnTh 346
497 GAACAGGAGGCCCTCTCAGGGGAGTGCAGAGATTGGAAACACAGACANTCA 546
346 rLysLeuLysThrLysValAspAspLeuGlnArgGluIleLysTyrMetL 363
547 GGCCCTGAAAGAGAAAGCAGATTCCTGAGTAAGGAAATCCAGTACTTCA 596
363 ysAsnLeuMetGluAspValCysLysAlaLys 373
597 AAGATCTGATTGAAGAGGTGAAGGAAGGCCAAA 628
```

seq_name: gb_est13:AA903759

seq_documentation_block:

LOCUS AA903759 849 bp mRNA EST 09-JUN-1998
DEFINITION OK66A04.s1 NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:151894 3' similar to gb:D90209 DNA-BINDING PROTEIN TAXREB67 (HUMAN);, mRNA sequence.
ACCESSION AA903759
VERSION AA903759.1 GI:3038882
KEYWORDS EST.

SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 849)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Feb 6, 1998 this sequence version replaced gi:2843818.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
Emmert-Buck, M.D., Ph.D.
CDNA Library Prepared by: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert Length: 1059 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 308.
Location/Qualifiers
1. .849
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1518894"
/clone_lib="NCI_CGAP_GC4"
/tissue_type="pooled germ cell tumors"
/lab_host="DH108"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from 3 pooled
germ cell tumors, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."
BASE COUNT 197 a 216 c 205 g 231 t
ORIGIN

alignment_scores:
Quality: 209.00 Length: 250
Ratio: 1.504 Gaps: 11
Percent Similarity: 55.600 Percent Identity: 30.000

alignment_block:
US-08-656-811A-1 x AA903759/rev ...
Align seg 1/1 to reverse of: AA903759 from: 1 to: 849
143 PheGluAsnHisLeuSerProGluGlnValAlaProVa 159
757 TTCGGGGAACTATTAGCGGCC...CCCAAAAGGGTAACCCCAAT 714
159 lileAsnLeuGluProValGluLeuThrAlaSerHisMetThrValles 176
713 T.....GCCCTT 706
176 erProAspGlyLeuLeuGlyMetGluLeuAlaSerGluSerLeuThr 192
705 CCCGAAGGTTTACAAACCCCGACAGGTGCC.....CCTTTACT 662
193 PheThrGluLeuAspPheValAsnPhaAsnAspSerAlaValGlySerIl 209
661 TTCCTAAAA...ACTTTTCCCTTTTCCAGGGGCTCTTGCTTCCACTC 615
209 eGlyGlyAlaGluLeu...LeuGlySerProLeuSerValAsp 224
614 CGGATCATCTCTTAGTTAGCTTGGGCGAGTGAAGTGGATATCACTCAA 565

225 ValGluSerThrIleSerPheSerGly..... 233
564 GGATAGGAAGCAGACTACACTGCTTACGTCCTGATCCCTCAGTG 515
234ProSerSerProGluThrSerGlnSerSerIleleGluS 247
514 CATAAAGGAGCAGACACCCCTTCAGATAATGATAGTTGCATCTGT...A 468
247 erSerProGluLeuThrLysValIleSerThrSerSerIleAspAlaSer 263
467 TGAGCCAGAGCTCTCTATCTG.....GGGTCT 442
264 LysArgPheSerProTyrSerArgSerSerLysSerLysGlnSerValLy 280
441 CCTCAGCACAGCCCTCTACAGGGGCTCTCCAATAGG...AGCCTCCC 395
280 sThrSerAspAlaLysAlaProArgLysThrArgThrProAlaGlnProV 297
394 ATCTCCAGGTCTTCTGTGGTCTGCCCGTCCCAAACTTACGATCCTC 345
297 alProGluHisValIleMet.....GluHisLeuAspLys 308
344 CTGAGAGAGAGATGCTAGCAGCAAAAGTAAAGGGTGAGAACTGGATAAG 295
309 LysAspArgLysLysLeuGlnAsnLysAsnAlaAlaIleArgTyrArgMe 325
294 AAGTGTAATAAATGGACCAAAACAGACAGCAGCCACTAGTACCGCCA 245
325 tLysLysLysGlyGluAlaGlnGlyIleLysGlyGluGlnGluLeuG 342
244 GAAGAAGGGCGGAGGAGGAGGCTCTTACTGGTGAAGCAAGAGCTGG 195
342 luGluLeuAsnThrLysLeuLysValAspAspLeuGlnArgGlu 358
194 AAAAGAAGACGAGGCTCTAAAGAGAGAGGGCGGATTCCTCGCCCAAGGAG 145
359 IleLysTyrMetLysAsnLeuMetGluAspValCysLysAlaLysGly 374
144 ATCCAGTACCTGAAGATTGATAGAGAGGTCCGCAAGGCAAGGGGG 97
seq_name: gb_est18:AI237099
seq_documentation_block:
LOCUS AI237099 545 bp mRNA EST 31-JAN-1999
DEFINITION EST233661 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
ROVDO64 3' end, mRNA sequence.
ACCESSION AI237099
VERSION AI237099.1 GI:3830605
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 545)
AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
TITLE Rat Genome Project: generation of a Rat EST (EST) Catalog & Rat
Gene Index
JOURNAL Unpublished (1998)
COMMENT On Apr 7, 1998 this sequence version replaced gi:3036733.
Other ESTs: TC61418
Contact: Lee, NH
ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: MJ3-21.
Location/Qualifiers
1. .545
source

/organism="Rattus sp."
 /db_xref="ATCC (inhost):2042258"
 /db_xref="taxon:10118"
 /clone="ROVDO64"
 /clone_lib="Normalized rat ovary, Bento Soares"
 /note="Organ: ovary; Vector: pT7T3Pac; Site_1: EcoRI;
 Site_2: NotI"

BASE COUNT 124 a 141 c 120 g 160 t
 ORIGIN

alignment_scores:
 Quality: 208.00 Length: 147
 Ratio: 2.237 Gaps: 3
 Percent Similarity: 63.265 Percent Identity: 38.776

alignment_block:
 US-08-656-811A-1 x AI237099/rev ..

Align seg 1/1 to reverse of: AI237099 from: 1 to: 545

```

242 SerSerIleGluSerProGluLeuTyLysValIleSerThr... 257
      ::::: ||| ::::: |||
539 GCTGCTTATATTACTCTAACCCCTCAGTGTGTAAGGAGGACACTCC 490
      ::::: ||| ::::: |||
258 .SerSerIleAspAlaSerLysArgPheSerProTyLysSerL 274
      ::::: ||| ::::: |||
489 CTCGTAGTGCACAGTGCATCTGTATGAGCCCTGAGTCTACCTGGGCT 440
      ::::: ||| ::::: |||
274 ySerLysGlnSerValLysThrSerAspAlaLys..... 285
      ::::: ||| ::::: |||
439 CTCGCCAACACAGCCCTTCCACCTCCAGGCCCCACCAGACAGTCTGCCT 390
      ::::: ||| ::::: |||
286 .....AlaProArgLysThrArg.....ThrProAl 294
      ::::: ||| ::::: |||
389 TCTCCAGGTGTTCTCGTGGTTCGACCCCAACCTTATGACCCACCTGG 340
      ::::: ||| ::::: |||
294 aGlnProValProGluHisValIleMetGluHisLeuAspLysAspA 311
      ::::: ||| ::::: |||
339 AGTAGTGTGACAGCTAAAGTGAAGACTGAAAGCTGGATGAAGACTGA 290
      ::::: ||| ::::: |||
311 rGlyLysLeuGlnAsnLysAsnAlaAlaIleArgTyArgMetLysLys 327
      ::::: ||| ::::: |||
289 AAAAGATGGAGCAAAACAAAGACAGCAGCTACTAGGTACCCGCAAGA 240
      ::::: ||| ::::: |||
328 LysGlyGluAlaGlnGlyLeuGlnGluGlnGluGluLe 344
      ::::: ||| ::::: |||
239 AGGCTGAGCAGGAGGCCCTCAGTGGGAGTGTAAAGAGCTAGAAAGAA 190
      ::::: ||| ::::: |||
344 uAsnThrLysLeuLysThrLysValAspAspLeuGlnArgGluLysT 361
      ::::: ||| ::::: |||
189 GAACGAGGCTCTGAAGAGAGAGGAGGATCTCTGCCCAAGAGATTTCAGT 140
      ::::: ||| ::::: |||
361 yrMetLysAsnLeuMetGluAspValCysLysAlaLysGly 374
      ::::: ||| ::::: |||
139 ATCTAAAGACCTGATAGAGAGGTCCTGTAAGGCAAGGGGG 99
      ::::: ||| ::::: |||

```

seq_name: gb_est20:AI408946

seq_documentation_block: 549 bp mRNA EST 09-FEB-1999
 LOCUS AI408946
 DEFINITION EST237237 Normalized rat kidney, Bento Soares Rattus sp. cDNA clone
 RKIDR39 3' end, mRNA sequence.
 ACCESSION AI408946
 VERSION AI408946
 KEYWORDS EST.
 SOURCE
 ORGANISM
 Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 549)
 AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,

TITLE
 JOURNAL
 COMMENT
 Kerlavage,A.R. and Adams,M.D.
 Rat Genome Project: Generation of a Rat EST (RESt) Catalog & Rat
 Gene Index
 Unpublished (1998)
 On Jan 19, 1998 this sequence version replaced gi:2286321.
 Contact: Lee, NH
 ATCC

The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@tigr.org
 Seq primer: M13-21.

FEATURES
 source

1..549
 /organism="Rattus sp."
 /db_xref="taxon:10118"
 /clone="RKIDR39"
 /clone_lib="Normalized rat kidney, Bento Soares"
 /note="Organ: kidney; Vector: pT7T3Pac; Site_1: EcoRI;
 Site_2: NotI"

BASE COUNT 126 a 142 c 121 g 160 t
 ORIGIN

alignment_scores:
 Quality: 208.00 Length: 147
 Ratio: 2.237 Gaps: 3
 Percent Similarity: 63.265 Percent Identity: 38.776

alignment_block:

US-08-656-811A-1 x AI408946/rev ..

Align seg 1/1 to reverse of: AI408946 from: 1 to: 549

```

242 SerSerIleGluSerProGluLeuTyLysValIleSerThr... 257
      ::::: ||| ::::: |||
539 GCTGCTTATATTACTCTAACCCCTCAGTGTGTAAGGAGGACACTCC 490
      ::::: ||| ::::: |||
258 .SerSerIleAspAlaSerLysArgPheSerProTyLysSerL 274
      ::::: ||| ::::: |||
489 CTCGTAGTGCACAGTGGCCTCTGTATGAGCCCTGAGTCTACCTGGGCT 440
      ::::: ||| ::::: |||
274 ySerLysGlnSerValLysThrSerAspAlaLys..... 285
      ::::: ||| ::::: |||
439 CTCGCCAACACAGCCCTTCCACCTCCAGGCCCCACCAGACAGTCTGCCT 390
      ::::: ||| ::::: |||
286 .....AlaProArgLysThrArg.....ThrProAl 294
      ::::: ||| ::::: |||
389 TCTCCAGGTGTTCTCGTGGTTCGACCCCAACCTTATGACCCACCTGG 340
      ::::: ||| ::::: |||
294 aGlnProValProGluHisValIleMetGluHisLeuAspLysAspA 311
      ::::: ||| ::::: |||
339 ACTTAGTGTGACAGCTAAAGTGAAGACTGAAAGTTGGATAAGAGCTGA 290
      ::::: ||| ::::: |||
311 rGlyLysLeuGlnAsnLysAsnAlaAlaIleArgTyArgMetLysLys 327
      ::::: ||| ::::: |||
289 AAAAGATGGAGCAAAACAAAGACAGCAGCTACTAGTACCCGCAAGAAG 240
      ::::: ||| ::::: |||
328 LysGlyGluAlaGlnGlyLeuGlnGluGlnGluGluLe 344
      ::::: ||| ::::: |||
239 AGGCTGAGCAGGAGGCCCTCAGTGGGAGTGTAAAGAGCTAGAAAGAA 190
      ::::: ||| ::::: |||
344 uAsnThrLysLeuLysThrLysValAspAspLeuGlnArgGluLysT 361
      ::::: ||| ::::: |||
189 GAACGAGGCTCTGAAGAGAGAGGAGGATCTCTGCCCAAGAGATTTCAGT 140
      ::::: ||| ::::: |||
361 yrMetLysAsnLeuMetGluAspValCysLysAlaLysGly 374
      ::::: ||| ::::: |||
139 ATCTAAAGACCTGATAGAGAGGTCCTGTAAGGCAAGGGGG 99
      ::::: ||| ::::: |||

```

seq_name: gb_est14:AA946061

```

seq_documentation_block: 551 bp mRNA EST 08-JAN-1999
LOCUS AA946061 EST201560 Normalized rat lung, Bento Soares Rattus sp. cDNA clone
DEFINITION RLUBA73 3' end, mRNA sequence.
ACCESSION AA946061
VERSION AA946061.1 GI:4132681
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 551)
AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
TITLE Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
Gene Index
JOURNAL Unpublished (1998)
COMMENT On May 1, 1998 this sequence version replaced gi:3105977.
Contact: Lee, NH
ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
FEATURES
Location/Qualifiers
source 1..551
/organism="Rattus sp."
/db_xref="taxon:10118"
/clone="RLUBA73"
/clone_lib="Normalized rat lung, Bento Soares"
/note="Organ: lung; Vector: pT7T3Pac; Site_1: EcoRI;
Site_2: NotI"
BASE COUNT 125 a 142 c 121 g 163 t
ORIGIN

alignment_scores:
Quality: 208.00 Length: 147
Ratio: 2.237 Gaps: 3
Percent Similarity: 63.265 Percent Identity: 38.776

alignment_block:
US-08-656-811a-1 x AA946061/rev ..
Align seg 1/1 to reverse of: AA946061 from: 1 to: 551

242 SerSerIleIleGluSerSerProGluLeuTyLysValIleSerThr.. 257
: : : : : ||| : : : : : |||
539 GCTGCTTATATTACTCTACCCCTCAGTGTGTAAGGAGGAGACACTCC 490
: : : : : ||| : : : : : |||

258 .SerSerIleAspAlaSerLysArgPheSerProTyTyrSerArgSerL 274
||| : : : : : ||| : : : : : |||
489 CTCGTGATGATGACAGTGCATCTGTATGAGCCCTGAGTCCCTACCTGGGCT 440
: : : : : ||| : : : : : |||

274 ySerLysGlnSerValLysThrSerAspAlaLys..... 285
: : : : : ||| : : : : : |||
439 CTCGCCAACACAGCCCTTCCACCTCCAGGCGCCACCACGAGCTGCCT 390
286 .....AlaProArgLysThrArg.....ThrProAl 294
: : : : : ||| : : : : : |||
389 TCCTCCAGGTGTTCTCTGTTCTCGACCCCAACCTTATGACCCACCTGG 340

294 aGlnProValProGluHisValIleMetGluHisLeuAspLysLysAsp 311
||| : : : : : ||| : : : : : |||
339 AGTTAGTGTGACAGCTAAAGTGAAGACTGAAAGCTTGGATAAGAAGCTGA 290
: : : : : ||| : : : : : |||

311 rgLysLysLeuGlnAsnLysAsnAlaIleArgTyArgMetLysLys 327
: : : : : ||| : : : : : |||
289 AAAAGATGGAGCAAAAGACAGCAGCTACTAGTACCGCCGCAAGAAG 240

```

```

328 LysGlyGluAlaGlnGlyIleLysGlyGluGluGlnGluLeuGluLe 344
: : : : : ||| : : : : : |||
239 AGGCTGAGCAGGAAGCCCTCACTGGCAGTGTAAAGAGCTAGAAAGAA 190
: : : : : ||| : : : : : |||

344 uAsnThrLysLysThrLysValAspAspLeuGlnArgGluIleLysT 361
||| : : : : : ||| : : : : : |||
189 GAACGAGGCTCTGAAGAGAGAGGAGATCTCTGCCCAAGAGATTCAGT 140
: : : : : ||| : : : : : |||

361 yrMetLysAsnLeuMetGluAspValCysLysAlaLysGly 374
||| : : : : : ||| : : : : : |||
139 ATCTAAGAGACTGATAGAGAGGTCCTTAAGGCAAGGGG 99
: : : : : ||| : : : : : |||

seq_name: gb_est18:AI237167
seq_documentation_block: 533 bp mRNA EST 31-JAN-1999
LOCUS AI237167 EST233729 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
DEFINITION ROVDP45 3' end, mRNA sequence.
ACCESSION AI237167
VERSION AI237167.1 GI:3830673
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 533)
AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
TITLE Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
Gene Index
JOURNAL Unpublished (1998)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2285377.
Other_ESTs: TC61418
Contact: Lee, NH
ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
FEATURES
Location/Qualifiers
source 1..533
/organism="Rattus sp."
/db_xref="ATCC (inhost):2042319"
/db_xref="taxon:10118"
/clone="ROVDP45"
/note="Normalized rat ovary, Bento Soares"
/note="Organ: ovary; Vector: pT7T3Pac; Site_1: EcoRI;
Site_2: NotI"
BASE COUNT 121 a 138 c 118 g 156 t
ORIGIN

alignment_scores:
Quality: 207.00 Length: 144
Ratio: 2.275 Gaps: 3
Percent Similarity: 63.194 Percent Identity: 39.583

alignment_block:
US-08-656-811a-1 x AI237167/rev ..
Align seg 1/1 to reverse of: AI237167 from: 1 to: 533

245 IleGluSerSerProGluLeuTyLysValIleSerThr...SerSerI 260
||| : : : : : ||| : : : : : |||
525 ATTACTTAACCCCTCAGTGTGTAAGGAGGAGACACACTCCCTCTGATAG 476
: : : : : ||| : : : : : |||

260 eAspAlaSerLysArgPheSerProTyTyrSerArgSerLysSerLys 277
||| : : : : : ||| : : : : : |||
475 TGACAGTGGCATCTGTATGATGACCTGAGTCTACCTGGCTCTCCCAAC 426

```

```
277 lnSerValLysThrSerAspAlaLys..... 285
285 ACAGCCCTTCCACCTCCAGGCCCCACACAGACAGTCTGCCCTTCCACAGGT 376
286 AlaProArgLysThrArg.....ThrProAlaGlnProVa 297
375 GTTCTCGTGGTTCTCGACCCCAACCTTATGACCCACCTGGAGTAGTGT 326
297 lProGluHisValIleMetGluHisLeuAspLysLysAspArgLysLysL 314
325 GACAGCTAAAGTGAAGACTGAAAGTTGGATAAAGCTGAAAGATGG 276
314 euGlnAsnLysAsnAlaAlaIleArgTyrArgMetLysLysGlyGlu 330
275 AGCAAAACAGACAGACAGCTACTAGGTACCGCCAGAGAGAGAGGCTGAG 226
331 AlaGlnGlyIleLysGlyGluGluGlnGluLeuGluLeuAsnThrLy 347
225 CAGGAAGCCCTCAGTGGCGAGTGTAAAGAGCTAGAAAAGAGAAGAGGCC 176
347 sLeuLysThrLysValAspAspLeuGlnArgGluIleLysTyrMetLys 364
175 TCTGAAAGAGAAGGACAGATTCTCTCGCCAAAGAGATTGATCTATAAG 126
364 snLeuMetGluAspValCysLysLysLysGly 374
125 ACCTGATAGAGAGGTCGCTAAGCAAGGGGG 94
```

seq_name: gb_est20:AI412959

seq_documentation_block: 540 bp mRNA EST 09-FEB-1999
LOCUS AI412959
DEFINITION EST241259 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
RBRED32 3' end. mRNA sequence.

ACCESSION AI412959
VERSION AI412959.1 GI:4256463
KEYWORDS EST.

SOURCE Rattus sp.
ORGANISM Rattus sp.

REFERENCE 1 (bases 1 to 540)
AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
TITLE Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat
Gene Index
JOURNAL Unpublished (1998)
COMMENT On Sep 25, 1998 this sequence version replaced gi:3658537.
Contact: Lee, NH

ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M3-21.
FEATURES
Location/Qualifiers
1..540
/organism="Rattus sp."
/db_xref="taxon:10118"
/clone="RBRED32"
/clone_lib="Normalized rat brain, Bento Soares"
/note="Organ: brain; Vector: pT7T3Pac; Site_1: EcoRI;
Site_2: NotI"
BASE COUNT 123 a 140 c 116 g 161 t
ORIGIN

alignment_scores:
Quality: 207.00 Length: 147
Ratio: 2.275 Gaps: 2
Percent Similarity: 61.905 Percent Identity: 37.415

alignment_block:
US-08-656-811A-1 x AI412959/rev ..

Align seg 1/1 to reverse of: AI412959 from: 1 to: 540

```
242 SerSerIleIleGluSerSerProGluLeuTyrLysValIleSerThr... 257
539 GCTGCTTATATTACTCTAAACCCCTCAGTGTGTAAAGGAGGAGACACTCC 490
258 SerSerIleAspAlaSerLysArgPheSerProTyrSerArgSerSerL 274
489 CTCGTAGTAGTACAGTGGCATCTGTATGAGCCCTGAGTCCTACATGGGCT 440
274 ysSerLysGlnSerValLysThrSerAspAla..... 284
439 CTCGCCAACACAGCCCTTCCACCTCCAGGCCCCACACAGACAGTCTGCCT 390
285 .....LysAlaProArgLysThrArgThrProAl 294
389 TCTCCAGGTGTTTCATCGTGGTTCGACCCCAACCTTATGACCCACCTGG 340
294 aglnProValProGluHisValIleMetGluHisLeuAspLysLysAspA 311
339 ACTTACTGTGACAGCTAAAGTGAAGACTCAAAAGTTGGATAAGAAGCTGA 290
311 rgLysLysLeuGlnAsnLysAsnAlaAlaIleArgTyrArgMetLysLys 327
289 AAAAGATGGAGCAAAACAGACAGCAGCTACTAGTACCGCCAGAGAAG 240
328 LysGlyGluAlaGlnGlyIleLysGlyGluGlnGluLeuGluGluLe 344
239 AGGCTGAGCAGGAGGAGCCCTCACTGCGAGTGTAAAGAGCTAGAAAAGAA 190
344 uasnThrLysLysLeuLysThrLysValAspLeuGlnArgGluIleLysT 361
189 GAACGAGGCTCTGAAAGAGAAGGAGGAGATTCTCTGCCCAAGAGATTCA 140
361 yrMetLysAsnLeuMetGluAspValCysLysLysAlaLysGly 374
139 ATCTAAAGACCTGATAGAGAGGTCCTGTAAGGCAAGGGGG 99
```

seq_name: gb_est23:AI649113

seq_documentation_block: 737 bp mRNA EST 30-APR-1999
LOCUS AI649113
DEFINITION uk25R03.x1 Sugano mouse embryo mewa Mus musculus cDNA clone
IMAGE:1970089.3' similar to gb:D90209 DNA-BINDING PROTEIN TAXREB67
(HUMAN); gb:M94087 M.musculus mATF4 (MOUSE);, mRNA sequence.

ACCESSION AI649113
VERSION AI649113.1 GI:4729947
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 737)

AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
TITLE The WashU-NCI Mouse EST Project 1999
JOURNAL Unpublished (1999)
COMMENT On Feb 18, 1999 this sequence version replaced gi:42999336.
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

MGT-986809

This clone was previously sequenced on the 5' end only, this new data is from the 3' end
Seq primer: custom primer used
High quality sequence stop: 513

FEATURES	SOURCE
1. Feature 1	Source 1
2. Feature 2	Source 2
3. Feature 3	Source 3
4. Feature 4	Source 4
5. Feature 5	Source 5
6. Feature 6	Source 6
7. Feature 7	Source 7
8. Feature 8	Source 8
9. Feature 9	Source 9
10. Feature 10	Source 10
11. Feature 11	Source 11
12. Feature 12	Source 12
13. Feature 13	Source 13
14. Feature 14	Source 14
15. Feature 15	Source 15
16. Feature 16	Source 16
17. Feature 17	Source 17
18. Feature 18	Source 18
19. Feature 19	Source 19
20. Feature 20	Source 20
21. Feature 21	Source 21
22. Feature 22	Source 22
23. Feature 23	Source 23
24. Feature 24	Source 24
25. Feature 25	Source 25
26. Feature 26	Source 26
27. Feature 27	Source 27
28. Feature 28	Source 28
29. Feature 29	Source 29
30. Feature 30	Source 30
31. Feature 31	Source 31
32. Feature 32	Source 32
33. Feature 33	Source 33
34. Feature 34	Source 34
35. Feature 35	Source 35
36. Feature 36	Source 36
37. Feature 37	Source 37
38. Feature 38	Source 38
39. Feature 39	Source 39
40. Feature 40	Source 40
41. Feature 41	Source 41
42. Feature 42	Source 42
43. Feature 43	Source 43
44. Feature 44	Source 44
45. Feature 45	Source 45
46. Feature 46	Source 46
47. Feature 47	Source 47
48. Feature 48	Source 48
49. Feature 49	Source 49
50. Feature 50	Source 50
51. Feature 51	Source 51
52. Feature 52	Source 52
53. Feature 53	Source 53
54. Feature 54	Source 54
55. Feature 55	Source 55
56. Feature 56	Source 56
57. Feature 57	Source 57
58. Feature 58	Source 58
59. Feature 59	Source 59
60. Feature 60	Source 60
61. Feature 61	Source 61
62. Feature 62	Source 62
63. Feature 63	Source 63
64. Feature 64	Source 64
65. Feature 65	Source 65
66. Feature 66	Source 66
67. Feature 67	Source 67
68. Feature 68	Source 68
69. Feature 69	Source 69
70. Feature 70	Source 70
71. Feature 71	Source 71
72. Feature 72	Source 72
73. Feature 73	Source 73
74. Feature 74	Source 74
75. Feature 75	Source 75
76. Feature 76	Source 76
77. Feature 77	Source 77
78. Feature 78	Source 78
79. Feature 79	Source 79
80. Feature 80	Source 80
81. Feature 81	Source 81
82. Feature 82	Source 82
83. Feature 83	Source 83
84. Feature 84	Source 84
85. Feature 85	Source 85
86. Feature 86	Source 86
87. Feature 87	Source 87
88. Feature 88	Source 88
89. Feature 89	Source 89
90. Feature 90	Source 90
91. Feature 91	Source 91
92. Feature 92	Source 92
93. Feature 93	Source 93
94. Feature 94	Source 94
95. Feature 95	Source 95
96. Feature 96	Source 96
97. Feature 97	Source 97
98. Feature 98	Source 98
99. Feature 99	Source 99
100. Feature 100	Source 100

1. /737/
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone_name="IMAGE:1970069"
/clone_lib="Sugano mouse embryo cDNA"
/dev_stage="embryo, 14 dpc"
/lab_host="Dh10B"
/note="vector: pME18S-FL3; Site1: Dr. Saito's pME18S-FL3; Site2: Dr. Iwano's pME18S-FL3 vector with an oligo(dT) primer [ATGTGGCTTGGTCTTAAAGTGGG]; 1st strand double-stranded cDNA was ligated to a [ATGTGGCTTAAAGTGGG], digested and cloned into the pME18S-FL3 vector [5', site1: CACCATCATG]. XhoI should be used to insert. Size selection was performed with a 1.5% agarose gel. Library constructed by Dr. Saito (University of Tokyo Institute of Medical Science) for sequencing: 5' end primer: CTTTGTGCTTAAAGTGGG and 3' end primer: CACCATCATGCTTAAAGTGGG."

BASE COUNT	171 a	189 c	175 g	201 t	1 others
ORIGIN	CGACCTGCAGCTCGAGCACA."				

alignment_scores:		
Quality	207.00	Length: 267
Ratio	1.656	Gaps: 9
Percent Similarity	46.816	Percent Identity: 28.839

alignment block:

US-08-656-811A-1 x AI649113/rev

Align seg 1/1 to reverse of: AI649113 from: 1 to: 737

109 SerLysAspIleLeuSerSerThrLeuGlnPheProThrGlnProValAs 125
 633 TCCAGGGGGTCTCTCTCCACCT
 125 nIleProLeuTyGAlaSerHisGlyAlaGluAspPheSerAlaGluThrC 142
 609CAGNAGCATCTCTTAGTTTAGAGCTAG 582
 142 luPheGluAsnHisLeuSerProProAsp...SerProGluGlnValAla 157
 581 GCAGTGAAGTTCATATCTCTGAAGGAGACGAGGACCTGAC..... 541
 158 ProValIleAsnLeuGluProValGluLeuThrAlaSerHisMetThrVa 174
 540TCTGCTGCTTACACTTACTCT 521
 174 lIleSerProAspGlyLeuLeuGlyGlyMetGluLeuAlaSerGluSerL 191
 520 AATCCCTCCA..... 511
 191 euThrPheThrGluLeuAspPheValAsnPheAsnAspSerAlaValGly 207
 510 ..TGTTGAAGGAGGAGACACACTCCCTCTGCACATGCAGCT.....GGC. 469
 208 SerIleGlyAlaGluGluLeuLeuLeuLeuSerProLeuSerValAspAs 224
 468 ATCTGTATGAGCCACAGACTCTACTGGGCTCTCCC..... 433
 224 pValGluSerThrIleSerPheSerGlyProSerSerProGluThrSerG 241
 432CAGCATGAGCCCTCCACTCTCA 411

```

241 lnSerSerIleIleGlnuSerSerProGluLeuTyLysValIleSerThr 257
   ::      ::      ::      ::      ::      ::      ::
410 GG.....GCCACCAGAC.....AATCTGCTTCTCCA 382

258 SerSerIleAspAlaSerLysArgPheSerProTyrSerArgSerSerly 274
   ::      ::      ::      ::      ::      ::      ::
381 GGTGGTTCCCGTGGGTCTCCCTCGGCCCAACCTTATGAC..... 343

274 sSerLysGlnSerValLysThrSerAspAlaLysAlaProArgLysThrA 291
   ::      ::      ::      ::      ::      ::      ::
343 ..... 343

291 rgThrProAlaGlnProValProGluHisValIleMetGluHisLeuAsp 307
   ::      ::      ::      ::      ::      ::      ::
342 .CCACTGGAGTTAGTTTGACACTAAAGTGAAGACTGAGAAATGGAT 295

308 LysLysAspArgLysLysLeuGlnAsnLysAsnAlaIleArgTyzAr 324
   ::      ::      ::      ::      ::      ::      ::
294 AAGAACTGAAAAAGATGGAGCAAAACAGACAGCAGCCACTAGGTACCG 245

324 gMetLysLysLysGlyGluAlaGlnGlyIleLysGlyGluGlnGluL 341
   ::      ::      ::      ::      ::      ::      ::
244 CCAGAAGAAGGGGCTGAGTAGAGAGGCCCTCACTGGCCAGTGTAAAGGAG 195

341 euGluGluLeuAsnThrLysLysLeuLysThrLysValAspAspleuGlnarg 357
   ::      ::      ::      ::      ::      ::      ::
194 TAGAAAAAAGAATGAGGCTCTGAAAGAGAAGGCGAGATCTCTGTGCCAAG 145

358 GluIleLysTyMetLysAsnLeuMetGluAspValCysLysAlaLysgl 374
   ::      ::      ::      ::      ::      ::      ::
144 GAGATCCAGTATCTGAAGAAGACCTGATAGAAGAGGTCCGTAAAGCGAAGGG 95

374 y 374
   |
94 G 94

```

seq name: qb est17:AT187106

```

seq_documentation_block: 548 bp mRNA EST 29-OCT-1998
LOCUS Ail187106 qe38d03.s1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
DEFINITION IMAGE:1741253 3' similar to gb:D90209 DNA-BINDING PROTEIN TAXREB67
(HUMAN);, mRNA sequence.
ACCESSION Ail187106
VERSION Ail187106.1 GI:3737744
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 548)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE Tumor Gene Index
UNPUBLISHED Unpublished (1997)
COMMENT On Apr 7, 1998 this sequence version replaced gi:3036753.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 748 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 398.
Location/Qualifiers
1. .548
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1741253"
/clone_lib="Soares_fetal_lung_NbHL19W"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: lung; Vector: p7T3D (Pharmacia) with a

```



```
133 GCATAGCCCTCCACCTCCAGG.....GCCCCACCAGAC. 166
251 euYrLysValIleSerThrSerSerIleAspAlaSerLysArgPheSer 267
   ::::: ||| ::::: ::::: ||| :::::
167 .....ATCTGCCTTCTCCAGGTGGTCCCGTGGGTCTCTCGGCCCAA 211
   :::::
268 ProTyrSerArgSerLysSerLysGlnSerValLysThrSerAspAl 284
   |||||:::
212 CTTATGAC..... 220
284 aLysAlaProArgLysThrArgThrProAlaGlnProValProGluHisV 301
   |||||:::
221 .....CCACCTGGAGTTAGTTTGACAGCTAAAG 248
301 alIleMetGluHisLeuAspLysLysArgLysLysLeuGlnAsnLys 317
   ||| ||||| ||||| ::::: |||||
249 TGAAGACTGGAANTTGGTAAGAGCTGAANAAGATGGACCAAAACAAG 298
318 AsnAlaIleArgTyrArgMetLysLysLysGlyGluAlaGlnGlyI 334
   ::||| ||||| ::||| ::||| ::||| ::||| ::|||
299 ACAGCAGCACTAGGTACCGCCAGAGAGCGGCTGAGCAGGAGGCCCT 348
334 eLysGlyGluGlnGluLeuGluGluLeuAsnThrLysLeuLysThrL 351
   : ||||| ::||| ::||| ::|||
349 CACTGGCGAGTGTAAAGAGGCTAGAAAAAAGAATGAGGCTCTGAAAGAGA 398
351 ysValAspLeuGlnArgGluIleLysTyrMetLysAspLeuMetGlu 367
   ||||| ::||| ::||| ::||| ::||| ::||| ::|||
399 AGCAGATTCTTCGCCAAGAGATCCAGTATCTGAAAGACCTGTATAGA 448
368 AspValCysLysAlaLysGly 374
   ::||| ::|||
449 GAGGTCCGTAAAGCAAGGGG 469
```